Appendix A

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:35:18 ; Search time 18 Seconds
(without alignments)
608.735 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	885	69.0	268	1	T4S9_HUMAN	O60628 homo sapien
2	324	25.3	294	1	TNE7_HUMAN	O95858 homo sapien
3	251	19.6	238	1	T4S7_MOUSE	Q9dck3 mus musculu
4	243	18.9	238	1	T4S7_HUMAN	O14817 homo sapien
5	242	18.9	253	1	C151_RAT	Q9qza6 rattus norv
6	240.5	18.7	245	1	T4S6_HUMAN	O43657 homo sapien
7	240	18.7	249	1	T4S2_HUMAN	P41732 homo sapien
8	239.5	18.7	245	1	T4S6_MOUSE	O70401 mus musculu
9	239	18.6	218	1	CD53_MOUSE	Q61451 mus musculu
10	239	18.6	253	1	C151_CERAE	Q9mym2 cercopithec
11	239	18.6	253	1	C151_MOUSE	O35566 mus musculu
12	236	18.4	239	1	TNE5_HUMAN	O75954 homo sapien
13	236	18.4	249	1	T4S2_MOUSE	Q62283 mus musculu
14	235.5	18.4	219	1	CD53_HUMAN	P19397 homo sapien
15	234	18.2	253	1	C151_HUMAN	P48509 homo sapien
16	233.5	18.2	218	1	CD53_RAT	P24485 rattus norv
17	231	18.0	237	1	T4S3_HUMAN	P19075 homo sapien
18	228	17.8	267	1	CD82_HUMAN	P27701 homo sapien
19	220	17.1	237	1	CD63_MOUSE	P41731 mus musculu
20	217	16.9	266	1	CD82_MOUSE	P40237 mus musculu
21	215.5	16.8	236	1	CD63_BOVIN	Q9xsk2 bos taurus
22	215.5	16.8	266	1	CD82_RAT	O70352 rattus norv
23	214.5	16.7	225	1	CD9_PIG	Q8wmq3 sus scrofa
24	214	16.7	237	1	CD63_HUMAN	P08962 homo sapien
25	214	16.7	237	1	CD63_RABIT	Q28709 oryctolagus
26	211	16.4	237	1	CD63_RAT	P28648 rattus norv
27	210	16.4	236	1	CD81_MOUSE	P35762 mus musculu
28	208	16.2	225	1	CD9_MOUSE	P40240 mus musculu
29	208	16.2	236	1	CD81_RAT	Q62745 rattus norv
30	208	16.2	236	1	CD81_SAGOE	Q9n0j9 saguinus oe
31	207	16.1	236	1	CD81_CERAE	O97703 cercopithec
32	207	16.1	236	1	CD81_HUMAN	P18582 homo sapien
33	206	16.1	225	1	CD9_RAT	P40241 rattus norv
34	202	15.7	227	1	CD9_CERAE	P30409 cercopithec
35	201	15.7	227	1	CD9_HUMAN	P21926 homo sapien
36	200.5	15.6	281	1	CD37_MOUSE	Q61470 mus musculu
37	200.5	15.6	281	1	CD37_RAT	P31053 rattus norv
38	199	15.5	225	1	CD9_FELCA	P40239 felis silve
39	197	15.4	221	1	TSN2_MOUSE	Q922j6 mus musculu
40	197	15.4	221	1	TSN2_RAT	Q9jjw1 rattus norv
41	195.5	15.2	218	1	IM23_SCHMA	P19331 schistosoma
42	190.5	14.8	218	1	IM23_SCHHA	Q26499 schistosoma
43	189	14.7	225	1	CD9_BOVIN	P30932 bos taurus
44	188.5	14.7	218	1	IM23_SCHJA	P27591 schistosoma
45	188	14.7	221	1	TSN2_HUMAN	O60636 homo sapien

ALIGNMENTS

RESULT 1
T4S9_HUMAN
ID T4S9_HUMAN STANDARD; PRT; 268 AA.
AC O60628; O60746; Q9JLY1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane 4 superfamily, member 8 (Tetraspanin 5) (Tspan-5)
DE (Tetraspan NET-4).
GN TM4SF9 OR TSPAN5.
OS Homo sapiens (Human), and
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606, 10090;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RX MEDLINE=98390278; PubMed=9714763;
RA Todd S.C., Doctor V.S., Levy S.;
RT "Sequences and expression of six new members of the tetraspanin/TM4SF
RT family.";
RL Biochim. Biophys. Acta 1399:101-104(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RA Rubinstein E., Serru V., Boucheix C.;
RT "New tetraspans identified in the EST database.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Human; TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse; STRAIN=Swiss Webster / NIH;

```
RA   Garcia-Frigola C., de Lecea L., Soriano E.;
RT   "Mouse Tspan-5 cDNA cloning.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF053455; AAC69712.1; -.
DR   EMBL; AF065389; AAC17120.1; -.
DR   EMBL; BC009704; AAH09704.1; -.
DR   EMBL; AF121344; AAF28869.1; -.
DR   PIR; A59261; A59261.
DR   MGD; MGI:1928096; Tm4sf9.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0008583; P:mystery cell fate differentiation (sensu Dr. . .; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; FALSE_NEG.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN       1     17      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     18    38      POTENTIAL.
FT   DOMAIN       39    61      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62    82      POTENTIAL.
FT   DOMAIN       83    92      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     93    113     POTENTIAL.
FT   DOMAIN       114   232     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     233   253     POTENTIAL.
FT   DOMAIN       254   268     CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     49    49      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     169   169     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     174   174     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     232   232     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     91    94      MISSING (IN REF. 1).
SQ   SEQUENCE   268 AA;  30337 MW;  7F4480BD0FA6192D CRC64;

  Query Match             69.0%;  Score 885;  DB 1;  Length 268;
  Best Local Similarity   79.9%;  Pred. No. 1.4e-67;
  Matches  155;  Conservative   16;  Mismatches   21;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| |||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| |||||||:|||||||||||||||| |||| |||| | :|||||||||||| | |||
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 178

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||| :|||
Db        179 RCGVPFSCCTKDPA 192


RESULT 2
TNE7_HUMAN
ID   TNE7_HUMAN     STANDARD;      PRT;   294 AA.
AC   O95858;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Tetraspan NET-7.
GN   TM4SF7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rubinstein E., Serru V., Dessen P., Boucheix C.;
RT   "New tetraspans identified in the EST database.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF120266; AAD17295.1; -.
DR   EMBL; BC003157; AAH03157.1; -.
DR   EMBL; BC004161; AAH04161.1; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005624; C:membrane fraction; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PROSITE; PS00421; TM4_1; FALSE_NEG.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN       1     23      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     24    44      POTENTIAL.
FT   DOMAIN       45    62      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     63    83      POTENTIAL.
FT   DOMAIN       84    93      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114     POTENTIAL.
FT   DOMAIN       115   235     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     236   256     POTENTIAL.
FT   DOMAIN       257   294     CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     118   118     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     189   189     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     230   230     N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   294 AA;  33165 MW;  71A6DC64D6CA6BAE CRC64;

  Query Match             25.3%;  Score 324;  DB 1;  Length 294;
  Best Local Similarity   33.9%;  Pred. No. 2.6e-20;
  Matches   64;  Conservative   45;  Mismatches   70;  Indels   10;  Gaps    3;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMSVL 77
              |: |  :: |||::||| |::|::|  |:     : :      | |  : :::| || ::
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:35:18 ; Search time 18 Seconds
(without alignments)
608.735 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	885	69.0	268	1	T4S9_HUMAN	O60628 homo sapien
2	324	25.3	294	1	TNE7_HUMAN	O95858 homo sapien
3	251	19.6	238	1	T4S7_MOUSE	Q9dck3 mus musculu
4	243	18.9	238	1	T4S7_HUMAN	O14817 homo sapien
5	242	18.9	253	1	C151_RAT	Q9qza6 rattus norv
6	240.5	18.7	245	1	T4S6_HUMAN	O43657 homo sapien
7	240	18.7	249	1	T4S2_HUMAN	P41732 homo sapien
8	239.5	18.7	245	1	T4S6_MOUSE	O70401 mus musculu
9	239	18.6	218	1	CD53_MOUSE	Q61451 mus musculu
10	239	18.6	253	1	C151_CERAE	Q9mym2 cercopithec
11	239	18.6	253	1	C151_MOUSE	O35566 mus musculu
12	236	18.4	239	1	TNE5_HUMAN	O75954 homo sapien
13	236	18.4	249	1	T4S2_MOUSE	Q62283 mus musculu
14	235.5	18.4	219	1	CD53_HUMAN	P19397 homo sapien
15	234	18.2	253	1	C151_HUMAN	P48509 homo sapien
16	233.5	18.2	218	1	CD53_RAT	P24485 rattus norv
17	231	18.0	237	1	T4S3_HUMAN	P19075 homo sapien
18	228	17.8	267	1	CD82_HUMAN	P27701 homo sapien
19	220	17.1	237	1	CD63_MOUSE	P41731 mus musculu
20	217	16.9	266	1	CD82_MOUSE	P40237 mus musculu
21	215.5	16.8	236	1	CD63_BOVIN	Q9xsk2 bos taurus
22	215.5	16.8	266	1	CD82_RAT	O70352 rattus norv
23	214.5	16.7	225	1	CD9_PIG	Q8wmq3 sus scrofa
24	214	16.7	237	1	CD63_HUMAN	P08962 homo sapien
25	214	16.7	237	1	CD63_RABIT	Q28709 oryctolagus
26	211	16.4	237	1	CD63_RAT	P28648 rattus norv
27	210	16.4	236	1	CD81_MOUSE	P35762 mus musculu
28	208	16.2	225	1	CD9_MOUSE	P40240 mus musculu
29	208	16.2	236	1	CD81_RAT	Q62745 rattus norv
30	208	16.2	236	1	CD81_SAGOE	Q9n0j9 saguinus oe
31	207	16.1	236	1	CD81_CERAE	O97703 cercopithec
32	207	16.1	236	1	CD81_HUMAN	P18582 homo sapien
33	206	16.1	225	1	CD9_RAT	P40241 rattus norv
34	202	15.7	227	1	CD9_CERAE	P30409 cerco
35	201	15.7	227	1	CD9_HUMAN	P21926 homo sa
36	200.5	15.6	281	1	CD37_MOUSE	Q61470 mus muscu
37	200.5	15.6	281	1	CD37_RAT	P31053 rattus norv
38	199	15.5	225	1	CD9_FELCA	P40239 felis silve
39	197	15.4	221	1	TSN2_MOUSE	Q922j6 mus musculu
40	197	15.4	221	1	TSN2_RAT	Q9jjw1 rattus norv
41	195.5	15.2	218	1	IM23_SCHMA	P19331 schistosoma
42	190.5	14.8	218	1	IM23_SCHHA	Q26499 schistosoma
43	189	14.7	225	1	CD9_BOVIN	P30932 bos taurus
44	188.5	14.7	218	1	IM23_SCHJA	P27591 schistosoma
45	188	14.7	221	1	TSN2_HUMAN	O60636 homo sapien

ALIGNMENTS

RESULT 1
T4S9_HUMAN
ID T4S9_HUMAN STANDARD; PRT; 268 AA.
AC O60628; O60746; Q9JLY1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane 4 superfamily, member 8 (Tetraspanin 5) (Tspan-5)
DE (Tetraspan NET-4).
GN TM4SF9 OR TSPAN5.
OS Homo sapiens (Human), and
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606, 10090;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RX MEDLINE=98390278; PubMed=9714763;
RA Todd S.C., Doctor V.S., Levy S.;
RT "Sequences and expression of six new members of the tetraspanin/TM4SF
RT family.";
RL Biochim. Biophys. Acta 1399:101-104(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RA Rubinstein E., Serru V., Boucheix C.;
RT "New tetraspans identified in the EST database.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Human; TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse; STRAIN=Swiss Webster / NIH;

```
RA   Garcia-Frigola C., de Lecea L., Soriano E.;
RT   "Mouse Tspan-5 cDNA cloning.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF053455; AAC69712.1; -.
DR   EMBL; AF065389; AAC17120.1; -.
DR   EMBL; BC009704; AAH09704.1; -.
DR   EMBL; AF121344; AAF28869.1; -.
DR   PIR; A59261; A59261.
DR   MGD; MGI:1928096; Tm4sf9.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0008583; P:mystery cell fate differentiation (sensu Dr. . .; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; FALSE_NEG.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN        1     17       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     18     38       POTENTIAL.
FT   DOMAIN       39     61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62     82       POTENTIAL.
FT   DOMAIN       83     92       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     93    113       POTENTIAL.
FT   DOMAIN      114    232       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    233    253       POTENTIAL.
FT   DOMAIN      254    268       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     49     49       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    232    232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     91     94       MISSING (IN REF. 1).
SQ   SEQUENCE   268 AA;  30337 MW;  7F4480BD0FA6192D CRC64;

  Query Match           69.0%;  Score 885;  DB 1;  Length 268;
  Best Local Similarity  79.9%;  Pred. No. 1.4e-67;
  Matches  155;  Conservative   16;  Mismatches   21;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| ||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||||||||:||||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| |||||||::|||||||||||||| |||| |||| | :|||||||||||| | |||
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 178

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||| :|||
Db        179 RCGVPFSCCTKDPA 192


RESULT 2
TNE7_HUMAN
ID   TNE7_HUMAN     STANDARD;      PRT;   294 AA.
AC   O95858;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Tetraspan NET-7.
GN   NET7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rubinstein E., Serru V., Dessen P., Boucheix C.;
RT   "New tetraspans identified in the EST database.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF120266; AAD17295.1; -.
DR   EMBL; BC003157; AAH03157.1; -.
DR   EMBL; BC004161; AAH04161.1; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005624; C:membrane fraction; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PROSITE; PS00421; TM4_1; FALSE_NEG.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN        1     23       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     24     44       POTENTIAL.
FT   DOMAIN       45     62       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     63     83       POTENTIAL.
FT   DOMAIN       84     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       POTENTIAL.
FT   DOMAIN      115    235       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    236    256       POTENTIAL.
FT   DOMAIN      257    294       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    118    118       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    189    189       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    230    230       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   294 AA;  33165 MW;  71A6DC64D6CA6BAE CRC64;

  Query Match           25.3%;  Score 324;  DB 1;  Length 294;
  Best Local Similarity  33.9%;  Pred. No. 2.6e-20;
  Matches   64;  Conservative   45;  Mismatches   70;  Indels   10;  Gaps    3;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMSVL 77
              |: |  :: |||::||| |::|::|  |:     : :      | |  : :::| || ::
```

```
Db         21 KFSLIIYSTVFWLIGALVLSVGIYAEVERQKYKTLES----AFLAPAIILILLGVVMFMV 76

Qy         78 GFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVKAYR 137

Db         77 SFIGVLASLRDNLYLLQAFMYILGICLIMELIGGVVALTFRNQTIDFLNDNIRRGIENYY 136

Qy        138 DDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAMSS 197

Db        137 DDLDFKNIMDFVQKKFKCCGGEDYRDWSKNQYHDCS--APGPLACGVPYTCCIR----NT 190

Qy        198 TPSVAMMSG 206

Db        191 TEVVNTMCG 199


RESULT 3
T4S7_MOUSE
ID   T4S7_MOUSE     STANDARD;      PRT;   238 AA.
AC   Q9DCK3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transmembrane 4 superfamily, member 7.
GN   TM4SF7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: FORMS A COMPLEX WITH INTEGRINS (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AK002709; BAB22301.1; -.
DR   EMBL; BC003482; AAH03482.1; -.
DR   MGD; MGI:1928097; Tm4sf7.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       POTENTIAL.
FT   DOMAIN       35     55       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     56     76       POTENTIAL.
FT   DOMAIN       77     85       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     86    106       POTENTIAL.
FT   DOMAIN      107    201       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    202    222       POTENTIAL.
FT   DOMAIN      223    238       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   238 AA;  26053 MW;  AA916EF6078777FA CRC64;

  Query Match             19.6%;  Score 251;  DB 1;  Length 238;
  Best Local Similarity   31.3%;  Pred. No. 3e-14;
  Matches   67;  Conservative   28;  Mismatches   95;  Indels   24;  Gaps    9;

Qy         14 GCCG--KYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVG 71

Db          4 GCLQGVKYLMFAFNLLFWLGGCGVLGVGIWLAATQGNFATLS--SSFPSLSAANLLIVTG 61

Qy         72 GVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINN 131

Db         62 TFVMAIGFVGCIGALKENKCLLLTFFVLLLLVFLLEATIAVLFFAYSDKIDSYAQQDLKK 121

Qy        132 NVKAY--RDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCC 189

Db        122 GLHLYGTQGNVGLTNAWSIIQTDFRCCGVSNYTDW-FEVY------NATR----VPDSCC 170

Qy        190 VRDPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVW 223

Db        171 -----LEFSDSCGLHEPGTW-WKSPCYETVKA-W 197


RESULT 4
T4S7_HUMAN
ID   T4S7_HUMAN     STANDARD;      PRT;   238 AA.
AC   O14817;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transmembrane 4 superfamily, member 7 (Novel antigen 2) (NAG-2)
DE   (Tetraspanin 4) (Tspan-4).
GN   TM4SF7 OR NAG2 OR TSPAN4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98030601; PubMed=9360996;
```

```
RA   Tachibana I., Bodorova J., Berditchevski F., Zutter M.M., Hemler M.E.;
RT   "NAG-2, a novel transmembrane-4 superfamily (TM4SF) protein that
RT   complexes with integrins and other TM4SF proteins.";
RL   J. Biol. Chem. 272:29181-29189(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98390278; PubMed=9714763;
RA   Todd S.C., Doctor V.S., Levy S.;
RT   "Sequences and expression of six new members of the tetraspanin/TM4SF
RT   family.";
RL   Biochim. Biophys. Acta 1399:101-104(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: FORMS A COMPLEX WITH INTEGRINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MULTIPLE TISSUES BUT IS ABSENT IN
CC       BRAIN, LYMPHOID CELLS, AND PLATELETS.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF022813; AAC51864.1; -.
DR   EMBL; AF054841; AAC69717.1; -.
DR   EMBL; BC000389; AAH00389.1; -.
DR   EMBL; BC019314; AAH19314.1; -.
DR   PIR; A59265; A59265.
DR   Genew; HGNC:11859; TM4SF7.
DR   MIM; 602644; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0006461; P:protein complex assembly; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       POTENTIAL.
FT   DOMAIN       35     55       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     56     76       POTENTIAL.
FT   DOMAIN       77     85       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     86    106       POTENTIAL.
FT   DOMAIN      107    201       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    202    222       POTENTIAL.
FT   DOMAIN      223    238       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   238 AA;  26118 MW;  A6B9A5633065A492 CRC64;

  Query Match             18.9%;  Score 243;  DB 1;  Length 238;
  Best Local Similarity   29.7%;  Pred. No. 1.4e-13;
  Matches   62;  Conservative   33;  Mismatches   90;  Indels   24;  Gaps    9;

Qy         18 KYFLFGFNIVFWVVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMSVL 77
              || :| ||::||: |   | :|:|    :|  : :|  :    |    | :: |  :  :
Db         10 KYLMFAFNLLFWLGGCGVLGVGIWLAATQGSFATLS--SSFPSLSAANLLIITGAFVMAI 67

Qy         78 GFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVKAY- 136
              || ||:||::||  ||  | : | |:| ||    || | : | |       :   :  |
Db         68 GFVGCLGAIKENKCLLLTFFLLLLLVFLLEATIAILFFAYTDKIDRYAQQDLKKGLHLYG 127

Qy        137 -RDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAM 195
               : :: | |     |  : |||     ||   :|      | :|    || |||    :
Db        128 TQGNVGLTNAWSIIQTDFRCCGVSNYTDW-FEVY------NATR----VPDSCC-----L 171

Qy        196 SSTPSVAMMSGSNWSWSSRAP-YTPKAVW 223
               : |  : :   | |  :|| |    ||
Db        172 EFSESCGLHAPGTW-W--KAPCYETVKVW 197


RESULT 5
C151_RAT
ID   C151_RAT       STANDARD;      PRT;   253 AA.
AC   Q9QZA6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Platelet endothelial tetraspan antigen 3 (CD151 antigen).
GN   CD151 OR PETA3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RX   MEDLINE=21539251; PubMed=11682256;
RA   Hua L.V., Green M., Wong A., Warsh J.J., Li P.P.;
RT   "Tetraspan protein CD151: a common target of mood stabilizing drugs?";
RL   Neuropsychopharmacology 25:729-736(2001).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF192547; AAF05763.2; -.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN        1     18       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     19     39       POTENTIAL.
FT   DOMAIN       40     57       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     58     78       POTENTIAL.
FT   DOMAIN       79     91       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     92    112       POTENTIAL.
FT   DOMAIN      113    221       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    222    242       POTENTIAL.
FT   DOMAIN      243    253       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   253 AA;  28355 MW;  51876AF31B4DCB2B CRC64;
```

```
  Query Match           18.9%;  Score 242;  DB 1;  Length 253;
  Best Local Similarity  32.3%;  Pred. No. 1.8e-13;
  Matches   62;  Conservative   28;  Mismatches   82;  Indels   20;  Gaps    7;

Qy          8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              | | :  |    | || || :|  ||: |   :|:|:|     |      || |      |
Db          4 FNEKKATCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASSTYLATA 61

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
              :: || | |: | |  ||    :|   ||: : : | :|| ||:  ||||:|:      |
Db         62 YILVVAGVVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGILAYVY----YQQ 117

Qy        125 LNFFINNN-----VKAYRDD--IDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNP 177
              ||  :  |     :| |       : | :|  |: : |||:    ||  : :    : :
Db        118 LNTELKENLKDTMIKRYHQSGHEGVTNAVDKLQQEFHCCGSNNSRDWRDSEWIRSGEAD- 176

Qy        178 SRERCGVPFSCC 189
              ||    || |||
Db        177 SRV---VPDSCC 185
```

```
RESULT 6
T4S6_HUMAN
ID   T4S6_HUMAN     STANDARD;      PRT;   245 AA.
AC   O43657;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transmembrane 4 superfamily, member 6 (Tetraspanin 6) (Tspan-6) (T245
DE   protein) (Tetraspanin TM4-D) (A15 homolog).
GN   TM4SF6 OR TSPAN6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98390278; PubMed=9714763;
RA   Todd S.C., Doctor V.S., Levy S.;
RT   "Sequences and expression of six new members of the tetraspanin/TM4SF
RT   family.";
RL   Biochim. Biophys. Acta 1399:101-104(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99000849; PubMed=9782095;
RA   Maeda K., Matsuhashi S., Hori K., Xin Z., Mukai T., Tabuchi K.,
RA   Egashira M., Niikawa N.;
RT   "Cloning and characterization of a novel human gene, TM4SF6, encoding
RT   a protein belonging to the transmembrane 4 superfamily, and mapped to
RT   Xq22.";
RL   Genomics 52:240-242(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Maeda K., Matsuhashi Y.;
RT   "The cDNA cloning of a novel gene A15 homologue which encords a new
RT   member of the transmembrane 4 superfamily.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Puls K.L., Ni J., Liu D., Morahan G., Wright M.D.;
RT   "The molecular characterization of four tetraspanins.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF053453; AAC69710.1; -.
DR   EMBL; AF043906; AAC64257.1; -.
DR   EMBL; U84895; AAD00560.1; -.
DR   EMBL; AF133426; AAF08365.1; -.
DR   EMBL; AL035608; -; NOT_ANNOTATED_CDS.
DR   EMBL; BC012389; AAH12389.1; -.
DR   PIR; A59258; A59258.
DR   Genew; HGNC:11858; TM4SF6.
DR   MIM; 300191; -.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Transmembrane; Glycoprotein; Polymorphism.
FT   DOMAIN        1     19       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     20     40       POTENTIAL.
FT   DOMAIN       41     59       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     60     80       POTENTIAL.
FT   DOMAIN       81     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       POTENTIAL.
FT   DOMAIN      115    208       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    209    229       POTENTIAL.
FT   DOMAIN      230    245       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     108    108       A -> T (IN dbSNP:1802288).
FT                                /FTId=VAR_014494.
SQ   SEQUENCE   245 AA;  27563 MW;  811FAB19C2805BE2 CRC64;
```

```
  Query Match           18.7%;  Score 240.5;  DB 1;  Length 245;
  Best Local Similarity  31.6%;  Pred. No. 2.3e-13;
  Matches   61;  Conservative   30;  Mismatches   83;  Indels   19;  Gaps    6;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              |    :  |   |  | |  |  :  :||: | : ||:|:|   |  | |  :| :
Db          1 MASPSRRLQTKPVITCFKSVLLIYTFIFWITGVILLAVGIWG---KVSLENYFSLLNEKA 57

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
                :  :: :  | |: :||  ||    | : ::|| :::|| |:| :||   |: |||:
Db         58 TNVPFVLIATGTVIILLGTFGCFATCRASAWMLKLYAMFLTLVFLVELVAAIVGFVFRHE 117

Qy        121 IRDQLNFFINNNVKA---YRDDIDLQ-NLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLN 176
              |:   | | ||  ||    |    | : : :|  |    |||     ||        || |
Db        118 IK---NSFKNNYEKALKQYNSTGDYRSHAVDKIQNTLHCCGVTDYRDW--------TDTN 166

Qy        177 PSRERCGVPFSCC 189
```

```
                |: | | |||
Db        167 YYSEK-GFPKSCC 178


RESULT 7
T4S2_HUMAN
ID   T4S2_HUMAN     STANDARD;      PRT;   249 AA.
AC   P41732; Q8WVG5; Q9UEY9;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Transmembrane 4 superfamily, member 2 (Cell surface glycoprotein A15)
DE   (T-cell acute lymphoblastic leukemia associated antigen 1) (TALLA-1)
DE   (Membrane component, X chromosome, surface marker 1) (CD231 antigen).
GN   TM4SF2 OR MXS1 OR A15.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=93131291; PubMed=8420826;
RA   Emi N., Kitaori K., Seto M., Ueda R., Saito H., Takahashi T.;
RT   "Isolation of a novel cDNA clone showing marked similarity to
RT   ME491/CD63 superfamily.";
RL   Immunogenetics 37:193-198(1993).
RN   [2]
RP   SEQUENCE OF 4-249 FROM N.A.
RC   TISSUE=Peripheral blood;
RX   MEDLINE=95286314; PubMed=7768645;
RA   Takagi S., Fujikawa K., Imai T., Fukuhara N., Fukudome K.,
RA   Minegishi M., Tsuchiya S., Konno T., Hinuma Y., Yoshie O.;
RT   "Identification of a highly specific surface marker of T-cell acute
RT   lymphoblastic leukemia and neuroblastoma as a new member of the
RT   transmembrane 4 superfamily.";
RL   Int. J. Cancer 61:706-715(1995).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANT XMLR HIS-172.
RX   MEDLINE=20120717; PubMed=10655063;
RA   Zemni R., Bienvenu T., Vinet M.C., Sefiani A., Carrie A., Billuart P.,
RA   McDonell N., Couvert P., Francis F., Chafey P., Fauchereau F.,
RA   Friocourt G., desPortes V., Cardona A., Frints S., Meindl A.,
RA   Brandau O., Ronce N., Moraine C., Bokhoven H.V., Ropers H.-H.,
RA   Sudbrak R., Kahn A., Fryns J.-P., Beldjord C., Chelly J.;
RT   "A new gene involved in X-linked mental retardation identified by
RT   analysis of an X;2 balanced translocation.";
RL   Nat. Genet. 24:167-170(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Wang H., Gao X., Huang Y., Han J.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: NOT SOLELY EXPRESSED IN T CELLS. EXPRESSED IN
CC       ACUTE MYELOCYTIC LEUKEMIA CELLS OF SOME PATIENTS.
CC   -!- DISEASE: Defects in TM4SF2 are the cause of a form of X-linked
CC       non-specific mental retardation (XLMR).
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   -!- DATABASE: NAME=PROW; NOTE=PROW 1:38-40(2000);
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/guide/501790534_g.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10653; BAA01501.1; ALT_INIT.
DR   EMBL; D29808; BAA06191.1; ALT_INIT.
DR   EMBL; AJ250562; CAB65594.1; -.
DR   EMBL; AJ250563; CAB65594.1; JOINED.
DR   EMBL; AJ250564; CAB65594.1; JOINED.
DR   EMBL; AJ250565; CAB65594.1; JOINED.
DR   EMBL; AJ250566; CAB65594.1; JOINED.
DR   EMBL; AJ250567; CAB65594.1; JOINED.
DR   EMBL; AJ250568; CAB65594.1; JOINED.
DR   EMBL; AB062057; BAB55825.1; -.
DR   EMBL; AB062057; BAB55824.1; -.
DR   EMBL; BC018036; AAH18036.1; -.
DR   Genew; HGNC:11854; TM4SF2.
DR   MIM; 300096; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0006487; P:N-linked glycosylation; TAS.
DR   GO; GO:0009405; P:pathogenesis; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Transmembrane; Antigen; Disease mutation.
FT   DOMAIN        1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     17     40       POTENTIAL.
FT   DOMAIN       41     56       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     57     75       POTENTIAL.
FT   DOMAIN       76     86       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     87    112       POTENTIAL.
FT   DOMAIN      113    213       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    214    234       POTENTIAL.
FT   DOMAIN      235    249       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    155    155       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    158    158       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    177    177       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     172    172       P -> H (IN MRX).
FT                                /FTId=VAR_009259.
FT   CONFLICT     53     53       E -> K (IN REF. 5).
FT   CONFLICT    127    127       A -> T (IN REF. 5).
FT   CONFLICT    136    136       E -> K (IN REF. 3).
SQ   SEQUENCE   249 AA;  27574 MW;  F2CF4517DB388173 CRC64;

  Query Match             18.7%;  Score 240;  DB 1;  Length 249;
  Best Local Similarity   28.2%;  Pred. No. 2.6e-13;
  Matches   50;  Conservative   34;  Mismatches   81;  Indels   12;  Gaps    2;

Qy         16 CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMS 75
              | |  |  :: |||: | : ||:|:|    |  |    :|      : :: :  |  :
Db         14 CLKTLLIIYSFVFWITGVILLAVGVWG---KLTLGTYISLIAENSTNAPYVLIGTGTTIV 70

Qy         76 VLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVKA 135
```

```
                | |  ||    | : ::|| :::|| |:|  ||  ||  |||:  |:|       : ::
Db         71 VFGLFGCFATCRGSPWMLKLYAMFLSLVFLAELVAGISGFVFRHEIKDTFLRTYTDAMQT 130

Qy        136 YRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRD 192
              |  : :    :|  |   |||| :   :|: : ||            |:| |||: :
Db        131 YNGNDERSRAVDHVQRSLSCCGVQNYTNWSTSPYF---------LEHGIPPSCCMNE 178


RESULT 8
T4S6_MOUSE
ID   T4S6_MOUSE     STANDARD;      PRT;   245 AA.
AC   O70401;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transmembrane 4 superfamily, member 6 (Tetraspanin 6) (Tspan-6).
GN   TM4SF6 OR TSPAN6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98390278; PubMed=9714763;
RA   Todd S.C., Doctor V.S., Levy S.;
RT   "Sequences and expression of six new members of the tetraspanin/TM4SF
RT   family.";
RL   Biochim. Biophys. Acta 1399:101-104(1998).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF053454; AAC69711.1; -.
DR   PIR; A59260; A59260.
DR   MGD; MGI:1926264; Tm4sf6.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; FALSE_NEG.
KW   Transmembrane; Glycoprotein.
FT   DOMAIN        1     19       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     20     40       POTENTIAL.
FT   DOMAIN       41     59       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     60     80       POTENTIAL.
FT   DOMAIN       81     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       POTENTIAL.
FT   DOMAIN      115    208       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    209    229       POTENTIAL.
FT   DOMAIN      230    245       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   245 AA;  27333 MW;  25F68411E3602A12 CRC64;

  Query Match             18.7%;  Score 239.5;  DB 1;  Length 245;
  Best Local Similarity   29.5%;  Pred. No. 2.8e-13;
  Matches   56;  Conservative   32;  Mismatches   89;  Indels   13;  Gaps    3;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              |    :  |   | | |  |  :  :||: | : ||:|:|    |  | |  :| :
Db          1 MASPSRRLQTKPVITCLKSVLLIYTFIFWITGVILLAVGIWG---KVSLENYFSLLNEKA 57

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
               :  :: :  | |: :||  ||    | : ::|| :::|| ||| :||   |: |||:
Db         58 TNVPFVLIGTGTVIILLGTFGCFATCRTSAWMLKLYAMFLTLIFLVELVAAIVGFVFRHE 117

Qy        121 IRDQLNFFINNNVKAYRDDIDLQN-LIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSR 179
              |::       | :| |    | ::  :|  |    |||     ||    |:: |


Db        118 IKNSFKSNYENALKEYNSTGDYRSEAVDKIQSTLHCCGVTNYGDWKGTNYYSET------ 171

Qy        180 ERCGVPFSCC 189
                 | | |||
Db        172 ---GFPKSCC 178


RESULT 9
CD53_MOUSE
ID   CD53_MOUSE     STANDARD;      PRT;   218 AA.
AC   Q61451; Q61721;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Leukocyte surface antigen CD53 (Cell surface glycoprotein CD53).
GN   CD53.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=93200067; PubMed=8452817;
RA   Wright M.D., Rochelle J.M., Tomlinson M.G., Seldin M.F.,
RA   Williams A.F.;
RT   "Gene structure, chromosomal localization, and protein sequence of
RT   mouse CD53 (Cd53): evidence that the transmembrane 4 superfamily
RT   arose by gene duplication.";
RL   Int. Immunol. 5:209-216(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION IN HEMATOPOIETIC
CC       CELLS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X97227; CAA65864.1; -.
DR   EMBL; Z16078; CAA78892.1; -.
DR   MGD; MGI:88341; Cd53.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Antigen; Transmembrane.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     11     31       POTENTIAL.
FT   DOMAIN       32     53       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     54     68       POTENTIAL.
FT   DOMAIN       69     79       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     80    105       POTENTIAL.
FT   DOMAIN      106    180       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    181    205       POTENTIAL.
FT   DOMAIN      206    218       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    118    118       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   218 AA;  24032 MW;  8CA592EADCE15E3D CRC64;

  Query Match             18.6%;  Score 239;  DB 1;  Length 218;
  Best Local Similarity   30.9%;  Pred. No. 2.8e-13;
  Matches   63;  Conservative   29;  Mismatches   72;  Indels   40;  Gaps    7;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
              || || ||::||| |   |  |::       |||  |:  || ||      : |:|| ::
Db          9 KYVLFIFNLLFWVCGCCILGFGIYFLVQNTYGVLFRNLPFLT-LGN-----ILVIVGSII 62
```

```
Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
Db         63 MVVAFLGCMGSIKENKCLLMSFFVLLLIILLAEVTIAILLFVYEQKLNTLVAEGLNDSIQ 122

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPA 194
Db        123 HYHSDNSTMKAWDFIQTQLQCCGVNGSSDWT----------------SGPPSSC------ 160

Qy        195 MSSTPSVAMMSG-----SNWSWSS 213
Db        161 ----PSGADVQGCYNKAKSWFHSN 180
```

RESULT 10
C151_CERAE
ID C151_CERAE STANDARD; PRT; 253 AA.
AC Q9MYM2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Platelet-endothelial tetraspan antigen 3 (PETA-3) (CD151 antigen).
GN CD151.
OS Cercopithecus aethiops (Green monkey) (Grivet), and
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534, 9544;
RN [1]
RP SEQUENCE FROM N.A.
RA Shanmukhappa K., Kapil S.;
RT "CD151/PETA-3, a tetraspanin molecule, interacts with the 3'
RT untranslated region and partial nucleoprotein gene of porcine
RT reproductive and respiratory syndrome virus RNA.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF275665; AAF90151.1; -.
DR EMBL; AF275666; AAF90152.1; -.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
DR PROSITE; PS00421; TM4_1; 1.
KW Glycoprotein; Transmembrane.

FT DOMAIN	1	18	CYTOPLASMIC (POTENTIAL).
FT TRANSMEM	19	39	POTENTIAL.
FT DOMAIN	40	57	EXTRACELLULAR (POTENTIAL).
FT TRANSMEM	58	78	POTENTIAL.
FT DOMAIN	79	91	CYTOPLASMIC (POTENTIAL).
FT TRANSMEM	92	112	POTENTIAL.
FT DOMAIN	113	221	EXTRACELLULAR (POTENTIAL).
FT TRANSMEM	222	242	POTENTIAL.
FT DOMAIN	243	253	CYTOPLASMIC (POTENTIAL).
FT CARBOHYD	159	159	N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 253 AA; 28438 MW; CAD2780B63F644A4 CRC64;

Query Match 18.6%; Score 239; DB 1; Length 253;
Best Local Similarity 31.3%; Pred. No. 3.2e-13;
Matches 61; Conservative 30; Mismatches 78; Indels 26; Gaps 7;

```
Qy          8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
Db          4 FNEKKTTCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASGTYLATA 61

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
Db         62 YILVVAGAVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGVLAYVY----YQQ 117

Qy        125 LNFFINNNVKAYRDDI----------DLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTD 174
Db        118 LNTELKENLK---DTMAKRYHQPGHEAVTSAVDQLQQEFHCCGSNNSQDWRDSEWIR--- 171

Qy        175 LNPSRERCGVPFSCC 189
Db        172 LREARGRV-VPDSCC 185
```

RESULT 11
C151_MOUSE
ID C151_MOUSE STANDARD; PRT; 253 AA.
AC O35566; O89118;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Platelet-endothelial tetraspan antigen 3 (PETA-3) (GP27) (Membrane
DE glycoprotein SFA-1) (CD151 antigen).
GN CD151 OR PETA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97438231; PubMed=9294006;
RA Hasegawa H., Watanabe H., Nomura T., Utsunomiya Y., Yanagisawa K.,
RA Fujita S.;
RT "Molecular cloning and expression of mouse homologue of SFA-1/PETA-3
RT (CD151), a member of the transmembrane 4 superfamily.";
RL Biochim. Biophys. Acta 1353:125-130(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv;
RX MEDLINE=98267146; PubMed=9602068;
RA Fitter S., Seldin M.F., Ashman L.K.;
RT "Characterisation of the mouse homologue of CD151 (PETA-3/SFA-1);
RT genomic structure, chromosomal localisation and identification of 2
RT novel splice forms.";
RL Biochim. Biophys. Acta 1398:75-85(1998).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; D89290; BAA22447.1; -.
DR EMBL; AF033620; AAC25952.1; -.
DR EMBL; U89772; AAC25976.1; -.
DR MGD; MGI:1096360; Cd151.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
DR PROSITE; PS00421; TM4_1; 1.
KW Glycoprotein; Transmembrane.

FT DOMAIN	1	18	CYTOPLASMIC (POTENTIAL).
FT TRANSMEM	19	39	POTENTIAL.
FT DOMAIN	40	57	EXTRACELLULAR (POTENTIAL).
FT TRANSMEM	58	78	POTENTIAL.
FT DOMAIN	79	91	CYTOPLASMIC (POTENTIAL).
FT TRANSMEM	92	112	POTENTIAL.
FT DOMAIN	113	221	EXTRACELLULAR (POTENTIAL).
FT TRANSMEM	222	242	POTENTIAL.

```
FT   DOMAIN      243    253       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    237    237       G -> S (IN REF. 1).
SQ   SEQUENCE   253 AA;  28246 MW;  AEBEEBCE2D765F1B CRC64;

  Query Match             18.6%;  Score 239;  DB 1;  Length 253;
  Best Local Similarity   32.3%;  Pred. No. 3.2e-13;
  Matches   62;  Conservative   28;  Mismatches   82;  Indels   20;  Gaps    7;

Qy          8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              | | :  |    | || || :|  ||: |   :|:|:|    |     || |      |
Db          4 FNEKKATCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASSTYLATA 61

Qy         65 WLPVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
              :: || | |: | |  ||     :|   ||: : : | :|| ||:  ||||:|:      |
Db         62 YILVVAGVVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGILAYVY----YQQ 117

Qy        125 LNFFINNN-----VKAYRDD--IDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNP 177
              ||  :  |     || |       : : :|  |: : |||:    ||  : :    : :
Db        118 LNTELKENLKDTMVKRYHQSGHEGVSSAVDKLQQEFHCCGSNNSQDWQDSEWIRSGEAD- 176

Qy        178 SRERCGVPFSCC 189
              ||    || |||
Db        177 SRV---VPDSCC 185


RESULT 12
TNE5_HUMAN
ID   TNE5_HUMAN     STANDARD;      PRT;   239 AA.
AC   O75954;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Tetraspan NET-5.
GN   NET5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20185353; PubMed=10719184;
RA   Serru V., Dessen P., Boucheix C., Rubinstein E.;
RT   "Sequence and expression of seven new tetraspans.";
RL   Biochim. Biophys. Acta 1478:159-163(2000).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF089749; AAC35859.1; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005624; C:membrane fraction; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       POTENTIAL.
FT   DOMAIN       35     55       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     56     76       POTENTIAL.
FT   DOMAIN       77     85       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     86    106       POTENTIAL.
FT   DOMAIN      107    203       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    204    224       POTENTIAL.
FT   DOMAIN      225    239       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   239 AA;  26779 MW;  DD7BA332BF6584EB CRC64;

  Query Match             18.4%;  Score 236;  DB 1;  Length 239;
  Best Local Similarity   28.0%;  Pred. No. 5.4e-13;
  Matches   61;  Conservative   30;  Mismatches   91;  Indels   36;  Gaps    6;

Qy         15 CCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVM 74
              || || :| ||::||: |   | :|:|    :|  :  |       |    | : :| ::
Db          7 CCLKYMMFLFNLIFWLCGCGLLGVGIWLSVSQGNFATFS--PSFPSLSAANLVIAIGTIV 64

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
              | || ||:||::||  ||  | : | :|   ||   || ||: | : :     :   :
Db         65 MVTGFLGCLGAIKENKCLLLSFFIVLLVILLAELILLILFFVYMDKVNENAKKDLKEGLL 124

Qy        135 AY--RDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRD 192
              |   ::: |:|  :  |    |||     ||            |      ||  ||:
Db        125 LYHTENNVGLKNAWNIIQAEMRCCGVTDYTDW----------YPVLGENTVPDRCCM-- 171

Qy        193 PAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGC 230
                           |    |   ||  :|   |:||
Db        172 -------------ENSQGCGRNATTP--LW---RTGC 190


RESULT 13
T4S2_MOUSE
ID   T4S2_MOUSE     STANDARD;      PRT;   249 AA.
AC   Q62283; O88429; Q9DBS3;
DT   01-NOV-1997 (Rel. 35, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transmembrane 4 superfamily, member 2 (Cell surface glycoprotein A15)
DE   (PE31) (TALLA homolog).
GN   TM4SF2 OR MXS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE OF 5-249 FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RA   Nagira M., Ishikawa K.-I., Fujikawa K., Takagi S., Yoshie O.;
RT   "Molecular cloning and expression of mouse PE31 (TALLA).";
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 5-249 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Spleen;
RX   MEDLINE=20083609; PubMed=10617319;
RA   Hosokawa Y., Ueyama E., Morikawa Y., Maeda Y., Seto M., Senba E.;
RT   "Molecular cloning of a cDNA encoding mouse A15, a member of the
RT   transmembrane 4 superfamily, and its preferential expression in brain
RT   neurons.";
RL   Neurosci. Res. 35:281-290(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
```

```
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL PROLIFERATION AND CELL MOTILITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26483; BAA05493.1; ALT_INIT.
DR   EMBL; AF052492; AAC34579.1; ALT_INIT.
DR   EMBL; AK004776; BAB23554.1; -.
DR   MGD; MGI:1298407; Tm4sf2.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Transmembrane.
FT   DOMAIN        1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     17     40       POTENTIAL.
FT   DOMAIN       41     56       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     57     75       POTENTIAL.
FT   DOMAIN       76     86       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     87    112       POTENTIAL.
FT   DOMAIN      113    213       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    214    234       POTENTIAL.
FT   DOMAIN      235    249       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    155    155       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    158    158       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    177    177       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    146    146       R -> P (IN REF. 2).
SQ   SEQUENCE   249 AA;  27544 MW;  04B3A24D16B5DCD3 CRC64;

  Query Match             18.4%;  Score 236;  DB 1;  Length 249;
  Best Local Similarity   28.8%;  Pred. No. 5.7e-13;
  Matches   51;  Conservative   35;  Mismatches   79;  Indels   12;  Gaps    3;

Qy         16 CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMS 75
              |  |   |  :: |||: |  : ||:|:|    |  |    :|      :  :: :  |  :
Db         14 CLKTLLIIYSFVFWITGVILLAVGVWG---KLTLGTYISLIAENSTNAPYVLIGTGTTIV 70

Qy         76 VLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVKA 135
              |  |  ||    | : ::|| :::|| |:|  ||  ||  |||:  |:|       : ::
Db         71 VFGLFGCFATCRGSPWMLKLYAMFLSLVFLAELVAGISGFVFRHEIKDTFLRTYTDAMQN 130

Qy        136 YRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRD 192
              |  : :    :|  |   |||| :   :|: : ||    |:      |:| |||: :
Db        131 YNGNDERSRAVDHVQRSLSCCGVQNYTNWSSSPYF----LDH-----GIPPSCCMNE 178


RESULT 14
CD53_HUMAN
ID   CD53_HUMAN     STANDARD;      PRT;   219 AA.
AC   P19397;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Leukocyte surface antigen CD53 (Cell surface glycoprotein CD53).
GN   CD53 OR MOX44.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91055810; PubMed=1700763;
RA   Angelisova P., Vlcek C., Stefanova I., Lipoldova M., Horejsi V.;
RT   "The human leucocyte surface antigen CD53 is a protein structurally
RT   similar to the CD37 and MRC OX-44 antigens.";
RL   Immunogenetics 32:281-285(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91079522; PubMed=2258620;
RA   Amiot M.;
RT   "Identification and analysis of cDNA clones encoding CD53. A pan-
RT   leukocyte antigen related to membrane transport proteins.";
RL   J. Immunol. 145:4322-4325(1990).
CC   -!- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION IN HEMATOPOIETIC
CC       CELLS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: B CELLS, MONOCYTES, MACROPHAGES, NEUTROPHILS,
CC       SINGLE (CD4 OR CD8) POSITIVE THYMOCYTES, PERIPHERAL T CELLS.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD53 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd53.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37033; AAA35663.1; -.
DR   EMBL; M60871; AAA51951.1; -.
DR   PIR; A37243; A37243.
DR   Genew; HGNC:1686; CD53.
DR   MIM; 151525; -.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Antigen; Transmembrane.
FT   DOMAIN        1     11       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     12     32       POTENTIAL.
FT   DOMAIN       33     54       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     55     69       POTENTIAL.
FT   DOMAIN       70     80       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     81    106       POTENTIAL.
FT   DOMAIN      107    181       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    182    206       POTENTIAL.
FT   DOMAIN      207    219       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    129    129       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   219 AA;  24341 MW;  48D4633DBC9110B6 CRC64;

  Query Match             18.4%;  Score 235.5;  DB 1;  Length 219;
  Best Local Similarity   31.0%;  Pred. No. 5.5e-13;
  Matches   54;  Conservative   31;  Mismatches   64;  Indels   25;  Gaps    5;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
              || || ||::||: |   |  |::       |||  |: :|| ||      :||:|| ::
Db         10 KYVLFFFNLLFWICGCCILGFGIYLLIHNNFGVLFHNLPSLT-LGN-----VFVIVGSII 63

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
              |: | ||:|:::||  ||  | : | :|   |:   || ||::  : : :   : :::
Db         64 MVVAFLGCMGSIKENKCLLMSFFILLLIILLAEVTLAILLFVYEQKLNEYVAKGLTDSIH 123

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSC 188
              |  |   :   |  | :  |||  | :||                  | | ||
Db        124 RYHSDNSTKAAWDSIQSFLQCCGINGTSDWT---------------SGPPASC 161


RESULT 15
C151_HUMAN
ID   C151_HUMAN     STANDARD;      PRT;   253 AA.
```

```
AC   P48509; Q14826; Q96TE3;
DT   01-FEB-1996 (Rel. 33, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Platelet-endothelial tetraspan antigen 3 (PETA-3) (GP27) (Membrane
DE   glycoprotein SFA-1) (CD151 antigen).
GN   CD151.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Platelet;
RX   MEDLINE=95359431; PubMed=7632941;
RA   Fitter S., Tetaz T.J., Berndt M.C., Ashman L.K.;
RT   "Molecular cloning of cDNA encoding a novel platelet-endothelial cell
RT   tetra-span antigen, PETA-3.";
RL   Blood 86:1348-1355(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96186759; PubMed=8627808;
RA   Hasegawa H., Utsunomiya Y., Kishimoto K., Yanagisawa K., Fujita S.;
RT   "SFA-1, a novel cellular gene induced by human T-cell leukemia virus
RT   type 1, is a member of the transmembrane 4 superfamily.";
RL   J. Virol. 70:3258-3263(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   PubMed=11181065;
RA   Whittock N.V., McLean W.H.I.;
RT   "Genomic organization, amplification, fine mapping, and intragenic
RT   polymorphisms of the human hemidesmosomal tetraspanin CD151 gene.";
RL   Biochem. Biophys. Res. Commun. 281:425-430(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: Interacts with integrins alpha3beta1, alpha5beta1,
CC       alpha3beta1 and alpha6beta4, with CD9 and CD181.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in a variety of tissues including
CC       vascular endothelium and epidermis.
CC   -!- INDUCTION: BY HTLV-1.
CC   -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD151 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd151.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14650; AAA87064.1; -.
DR   EMBL; D29963; BAA06229.1; -.
DR   EMBL; AF315942; AAK14179.1; -.
DR   EMBL; BC001374; AAH01374.1; -.
DR   EMBL; BC013302; AAH13302.1; -.
DR   Genew; HGNC:1630; CD151.
DR   MIM; 602243; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005624; C:membrane fraction; TAS.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
DR   PROSITE; PS00421; TM4_1; 1.
KW   Glycoprotein; Transmembrane; Polymorphism.
FT   DOMAIN        1     18       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     19     39       POTENTIAL.
FT   DOMAIN       40     57       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     58     78       POTENTIAL.
FT   DOMAIN       79     91       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     92    112       POTENTIAL.
FT   DOMAIN      113    221       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    222    242       POTENTIAL.
FT   DOMAIN      243    253       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     132    132       R -> K.
FT                                /FTId=VAR_012490.
FT   VARIANT     137    137       S -> P.
FT                                /FTId=VAR_012491.
SQ   SEQUENCE   253 AA;  28313 MW;  5C81D7D62D750EAF CRC64;

  Query Match             18.2%;  Score 234;  DB 1;  Length 253;
  Best Local Similarity   31.0%;  Pred. No. 8.5e-13;
  Matches   61;  Conservative   29;  Mismatches   77;  Indels   30;  Gaps    7;

Qy          8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              | | :  |    | || || :|  ||: |   :|:|:|    |     || |     |
Db          4 FNEKKTTCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASGTYLATA 61

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
              :: || | |: | |  ||    :|   ||: : : | :|| ||:  ||||: :      |
Db         62 YILVVAGTVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGILAYAY----YQQ 117

Qy        125 LNFFINNNVKAYRDDID----------LQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTD 174
              ||  :  |:|   | :           : : :|  |: : |||:    ||  : :
Db        118 LNTELKENLK---DTMTRRYHQSGHEAVTSAVDQLQQEFHCCGSNNSQDWRDSEWIR--- 171

Qy        175 LNPSRERCG--VPFSCC 189
                |:|  |  || |||
Db        172 ---SQEAGGRVVPDSCC 185


Search completed: November 21, 2003, 13:44:05
Job time : 19 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 17:14:19 ; Search time 148 Seconds
(without alignments)
7569.124 Million cell updates/sec

Title: US-09-972-970-2
Perfect score: 2538
Sequence: 1 ccacgcgtccggccgcagcc.....aaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	234	9.2	1932	4	US-09-482-273-41	Sequence 41, Appl
	2	95	3.7	399	3	US-09-188-930-62	Sequence 62, Appl
	3	95	3.7	399	3	US-09-188-930-243	Sequence 243, App
	4	95	3.7	399	4	US-09-312-283C-62	Sequence 62, Appl
	5	95	3.7	399	4	US-09-312-283C-243	Sequence 243, App
	6	67.8	2.7	1624	3	US-08-430-225A-19	Sequence 19, Appl
	7	58.4	2.3	473	4	US-09-702-705-1565	Sequence 1565, Ap
	8	58.4	2.3	473	4	US-09-736-457-1565	Sequence 1565, Ap
	9	57.8	2.3	1151	2	US-08-807-044-2	Sequence 2, Appli
	10	57.2	2.3	827	3	US-09-333-599-5	Sequence 5, Appli
	11	57.2	2.3	827	4	US-09-499-781-5	Sequence 5, Appli
	12	57.2	2.3	870	3	US-09-333-599-1	Sequence 1, Appli
	13	57.2	2.3	870	4	US-09-499-781-1	Sequence 1, Appli
	14	57.2	2.3	1344	3	US-08-705-771-8	Sequence 8, Appli
	15	54.8	2.2	977	2	US-08-855-140-2	Sequence 2, Appli
	16	54.8	2.2	977	4	US-09-016-434-938	Sequence 938, App
	17	54.2	2.1	1452	2	US-08-807-044-4	Sequence 4, Appli
	18	54.2	2.1	1452	5	PCT-US91-04986-1	Sequence 1, Appli
	19	52.2	2.1	687	1	US-08-254-493-2	Sequence 2, Appli
	20	52.2	2.1	687	1	US-08-408-222B-2	Sequence 2, Appli
	21	52.2	2.1	1120	1	US-08-254-493-3	Sequence 3, Appli
	22	52.2	2.1	1120	1	US-08-408-222B-3	Sequence 3, Appli
	23	46.8	1.8	560	3	US-09-221-298-12	Sequence 12, Appl
c	24	46.2	1.8	371	4	US-09-643-597-292	Sequence 292, App
c	25	46.2	1.8	371	4	US-09-480-884A-292	Sequence 292, App
c	26	46.2	1.8	371	4	US-09-542-615A-292	Sequence 292, App
c	27	46.2	1.8	371	4	US-09-606-421B-292	Sequence 292, App
	28	44.4	1.7	1637	3	US-08-852-824-3	Sequence 3, Appli
c	29	43.6	1.7	390	3	US-09-197-649-7	Sequence 7, Appli
	30	43.2	1.7	639	4	US-09-252-991A-10430	Sequence 10430, A
	31	43.2	1.7	744	4	US-09-252-991A-10328	Sequence 10328, A
c	32	43.2	1.7	870	4	US-09-252-991A-10645	Sequence 10645, A
	33	43.2	1.7	1227	4	US-09-252-991A-10523	Sequence 10523, A
c	34	42.8	1.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	35	42.8	1.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	36	42.4	1.7	289	3	US-09-007-005-17	Sequence 17, Appl
	37	42.4	1.7	289	3	US-09-244-796-17	Sequence 17, Appl
c	38	42	1.7	801	3	US-09-020-956-16	Sequence 16, Appl
c	39	42	1.7	801	3	US-09-030-607-16	Sequence 16, Appl
c	40	42	1.7	801	4	US-09-439-313-16	Sequence 16, Appl
c	41	42	1.7	801	4	US-09-352-616A-16	Sequence 16, Appl
c	42	42	1.7	801	4	US-09-232-149A-16	Sequence 16, Appl
	43	41.4	1.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	44	41	1.6	711	4	US-09-252-991A-5727	Sequence 5727, Ap
c	45	41	1.6	1548	4	US-09-252-991A-5713	Sequence 5713, Ap

ALIGNMENTS

RESULT 1
US-09-482-273-41
; Sequence 41, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: PZ030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1022)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-41

Query Match 9.2%; Score 234; DB 4; Length 1932;
Best Local Similarity 58.8%; Pred. No. 1.4e-51;
Matches 448; Conservative 0; Mismatches 298; Indels 16; Gaps 2;

```
Qy      150 CCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGC 209
            ||   ||| ||       | ||||||| | || ||   |||||||||  | |||||||||
Db      146 CCTTCGTCAGCCCGCTGGTGAAATACCTGCTCTTCTTCTTCAACATGCTCTTCTGGGTGA 205

Qy      210 TGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCA 269
            |         || |  ||||  | ||  |||  ||  ||      ||||  ||       |
Db      206 TTTCCATGGTGATGGTGGCTGTGGGTGTCTACGCTCGGCTAATGAAGCATGCAGAAGCAG 265

Qy      270 ACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTG 329
              | |  |    ||| | |              | |||||||    ||| |  | ||||| |
Db      266 CCCTAGCCTGCCTGGCAG-----------TGGACCCTGCCATCCTGCTGATCGTGGTGG 313
```

```
Qy        330 GAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTT 389
Db        314 GTGTCCTCATGTTCCTGCTCACCTTCTGTGGCTGCATTGGGTCCCTCCGCGAGAACATCT 373

Qy        390 TCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAG 449
Db        374 GCCTCCTGCAGACGTTCTCCCTCTGCCTCACCGCTGTGTTCCTGCTGCAGCTGGCCGCTG 433

Qy        450 GGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACA 509
Db        434 GGATCCTGGGCTTCGTCTTCTCAGACAAGGCTCGAGGGAAAGTGAGTGAGATCATCAACA 493

Qy        510 ACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGG 569
Db        494 ATGCCATTGTGCACTACCGAGATGACTTGGATCTGCAGAACCTCATTGATTTTGGCCAGA 553

Qy        570 AATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCA 629
Db        554 AAAAGTTTAGCTGCTGTGGAGGGATTTCCTACAAGGACTGGTCTCAGAACATGTATTTCA 613

Qy        630 ACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCA 689
Db        614 ACTGCTCAGAAGACAACCCCAGTCGAGAGCGCTGCTCTGTGCCTTACTCCTGTTGCTTGC 673

Qy        690 GGGACCCTG----CGATGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGG 745
Db        674 CTACTCCTGACCAGGCAGTGATCAACACTATGTGTGGCCAAGGTATGCAGGCCTTTGACT 733

Qy        746 AGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGC 805
Db        734 ACTTGGAAGCTAGCAAAGTCATCTACACCAATGGCTGTATTGACAAGTTGGTCAACTGGA 793

Qy        806 TGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCT 865
Db        794 TACACAGCAACCTATTCTTACTTGGTGGTGTGGCTCTAGGCCTGGCCATCCCCCAGCTGG 853

Qy        866 TTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGG 907
Db        854 TGGGAATTCTGCTGTCCCAGATCCTAGTGAATCAGATCAAAG 895
```

RESULT 2
US-09-188-930-62
; Sequence 62, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 62
; LENGTH: 399
; TYPE: DNA
; ORGANISM: mouse
US-09-188-930-62

Query Match 3.7%; Score 95; DB 3; Length 399;
Best Local Similarity 68.6%; Pred. No. 2.1e-15;
Matches 131; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy        144 AGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCT 203
Db          2 AGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTT 61

Qy        204 GGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTC 263
```

```
Db         62 GGTTTTTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCC 121

Qy        264 TCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAG 323
Db        122 TCTCCAACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCT 181

Qy        324 TGGTTGGAGGC 334
Db        182 GAGTGGCCAGC 192
```

RESULT 3
US-09-188-930-243
; Sequence 243, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Mouse
US-09-188-930-243

Query Match 3.7%; Score 95; DB 3; Length 399;
Best Local Similarity 68.6%; Pred. No. 2.1e-15;
Matches 131; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Q         144 AGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCT 203
Db          2 AGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTT 61

Qy        204 GGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTC 263
Db         62 GGTTTTTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCC 121

Qy        264 TCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAG 323
Db        122 TCTCCAACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCT 181

Qy        324 TGGTTGGAGGC 334
Db        182 GAGTGGCCAGC 192
```

RESULT 4
US-09-312-283C-62
; Sequence 62, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 62
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-312-283C-62

  Query Match             3.7%;  Score 95;  DB 4;  Length 399;
  Best Local Similarity   68.6%;  Pred. No. 2.1e-15;
  Matches  131;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        144 AGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCT 203
              |||  || || ||| | || |||   ||||||||| | ||||||||||  |  | || |
Db          2 AGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTT 61

Qy        204 GGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTC 263
              || |  |||||     ||| || |  ||||| || ||||| |||  ||| || |||||
Db         62 GGTTTTTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCC 121

Qy        264 TCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAG 323
              |||||||||||||  |  | ||||| || || ||| | ||||| ||||||||  ||  |
Db        122 TCTCCAACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCT 181

Qy        324 TGGTTGGAGGC 334
                || |   ||
Db        182 GAGTGGCCAGC 192


RESULT 5
US-09-312-283C-243
; Sequence 243, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 243
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-312-283C-243

  Query Match             3.7%;  Score 95;  DB 4;  Length 399;
  Best Local Similarity   68.6%;  Pred. No. 2.1e-15;
  Matches  131;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        144 AGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCT 203
              |||  || || ||| | || |||   ||||||||| | ||||||||||  |  | || |
Db          2 AGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTT 61

Qy        204 GGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTC 263
              || |  |||||     ||| || |  ||||| || ||||| |||  ||| || |||||
Db         62 GGTTTTTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCC 121

Qy        264 TCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAG 323
              |||||||||||||  |  | ||||| || || ||| | ||||| ||||||||  ||  |
Db        122 TCTCCAACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCT 181

Qy        324 TGGTTGGAGGC 334
                || |   ||
Db        182 GAGTGGCCAGC 192


RESULT 6
US-08-430-225A-19
; Sequence 19, Application US/08430225A
; Patent No. 6204000
;  GENERAL INFORMATION:
;    APPLICANT:  Dong, Jin-Tang; Barrett,
;    APPLICANT:  J. Carl; Lamb, Patricia W.; Isaacs, John T.
;    TITLE OF INVENTION:  DIAGNOSTIC METHODS AND
;    TITLE OF INVENTION:  GENE THERAPY USING REAGENTS DERIVED FROM THE
;    TITLE OF INVENTION:  HUMAN METASTASIS SUPPRESSOR GENE KAI1
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,225A
;      FILING DATE:  28-APR-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RICHARD W. BORK
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4172
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1624 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-430-225A-19

  Query Match             2.7%;  Score 67.8;  DB 3;  Length 1624;
  Best Local Similarity   51.1%;  Pred. No. 5.8e-08;
  Matches  159;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy        170 AAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCC 229
              |||||||| || ||   ||||||| |  | ||||   | ||||| ||  || |||||| |
Db        193 AAATACTTTCTCTTCCTCTTCAACTTGATCTTCTTTATCCTGGGCGCAGTGATCCTGGGC 252

Qy        230 ATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGAT 289
              ||||  | |||  |  ||  || ||| |   | ||  |   ||          ||
Db        253 TTCGGGGTGTGGATCCTGGCCGACAAGAGCAGTTTCATCTCTGTCCTGCAAACCTCCTCC 312

Qy        290 CTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTG 349
                                  | |   | ||  | |  || || |  ||||       | ||
Db        313 AGCTCGCTTAGGATGGGGGCCTATGTCTTCATCGGCGTGGGGGCAGTCACTATGCTCATG 372

Qy        350 GGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCA 409
              |||||     |||||||||||| ||   ||    ||| |   | ||||||    | | | ||
Db        373 GGCTTCCTGGGCTGCATCGGCGCCGTCAACGAGGTCCGCTGCCTGCTGGGGCTGTACTTT 432

Qy        410 GTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTC 469
              |  |||||   ||| ||| || |      || || |  | |||  | |   || | |   |
Db        433 GCTTTCCTGCTCCTGATCCTCATTGCCCAGGTGACGGCCGGGGCCCTCTTCTACTTCAAC 492

Qy        470 AAGGACTGGAT 480
              | || |  | |
Db        493 ATGGGCAAGCT 503


RESULT 7
```

```
US-09-702-705-1565
; Sequence 1565, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C14
;  CURRENT APPLICATION NUMBER: US/09/702,705
;  CURRENT FILING DATE:  2000-10-30
;  NUMBER OF SEQ ID NOS:  1833
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1565
;   LENGTH: 473
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(473)
;   OTHER INFORMATION: n = A,T,C or G
US-09-702-705-1565

  Query Match             2.3%;  Score 58.4;  DB 4;  Length 473;
  Best Local Similarity   50.6%;  Pred. No. 9.3e-06;
  Matches  137;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy        321 TAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGG 380
              | |  || || | | || | | | || ||| |||| |||||| |||||   |  ||
Db         66 TCGCAGTGGGTGTCTTCCTCTTCCTGGTGGCTTTTGTGGGCTGCTGCGGGGCCTGCAAGG 125

Qy        381 AGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGC 440
              |||||  ||  ||  | |    ||  |  | || ||    || ||| | ||  |||||
Db        126 AGAACTATTGTCTTATGATCACGTTTGCCATCTTTCTGTCTCTTATCATGTTGGTGGAGG 185

Qy        441 TGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCT 500
              ||||  |||  ||    | ||  || || |  ||   | |        || | |||  |
Db        186 TGGCCGCAGCCATTGCTGGCTATGTGTTNAGAGATAAGGTGATGTCAGAGTTTAATAACA 245

Qy        501 TCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACT 560
              | |      |  |  |  ||   || | |  |||   ||          || | |||
Db        246 ACTTCCGGCAGCAGATGGAGAATTACCCGAAAAACAACCACACTGNTTCNATCCTGGACA 305

Qy        561 TTGCTCAGGAATATTGGTCTTGCTGCGGAGC 591
                  |||| | |||     ||||| || ||
Db        306 GGATGCAGGCAGATTTTAAGTGCTGTGGGGC 336


RESULT 8
US-09-736-457-1565
; Sequence 1565, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1565
;   LENGTH: 473
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(473)
;   OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1565

  Query Match             2.3%;  Score 58.4;  DB 4;  Length 473;
  Best Local Similarity   50.6%;  Pred. No. 9.3e-06;
  Matches  137;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy        321 TAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGG 380
              | |  || || | | || | | | || ||| |||| |||||| |||||   |  ||
Db         66 TCGCAGTGGGTGTCTTCCTCTTCCTGGTGGCTTTTGTGGGCTGCTGCGGGGCCTGCAAGG 125

Qy        381 AGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGC 440
              |||||  ||  ||  | |    ||  |  | || ||    || ||| | ||  |||||
Db        126 AGAACTATTGTCTTATGATCACGTTTGCCATCTTTCTGTCTCTTATCATGTTGGTGGAGG 185

Qy        441 TGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCT 500
              ||||  |||  ||    | ||  || || |  ||   | |        || | |||  |
Db        186 TGGCCGCAGCCATTGCTGGCTATGTGTTNAGAGATAAGGTGATGTCAGAGTTTAATAACA 245

Qy        501 TCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACT 560
              | |      |  |  |  ||   || | |  |||   ||          || | |||
Db        246 ACTTCCGGCAGCAGATGGAGAATTACCCGAAAAACAACCACACTGNTTCNATCCTGGACA 305

Qy        561 TTGCTCAGGAATATTGGTCTTGCTGCGGAGC 591
                  |||| | |||     ||||| || ||
Db        306 GGATGCAGGCAGATTTTAAGTGCTGTGGGGC 336


RESULT 9
US-08-807-044-2
; Sequence 2, Application US/08807044
; Patent No. 5863735
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/807,044
;      FILING DATE:  Herewith
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0224 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
```

```
;       TELEFAX:  415-845-4166
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1151 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;       LIBRARY:  SCORNOT01
;       CLONE:  663655
US-08-807-044-2

  Query Match             2.3%;  Score 57.8;  DB 2;  Length 1151;
  Best Local Similarity   51.3%;  Pred. No. 2.1e-05;
  Matches  162;  Conservative    0;  Mismatches  148;  Indels    6;  Gaps    1;

Qy        166 CGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCT 225
              ||  || ||| || |||| | ||||||| |  | |||||| ||   ||          | ||
Db        237 CGTCAAGTACCTCATGTTCGCCTTCAACCTGCTNTTCTGGCTGGGAGGCTGTGGCGTGCT 296

Qy        226 GGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGAC 285
              ||   ||||| |||||      |      ||||     || |||   | ||         |
Db        297 GGGTGTCGGCATCTGGCTGGCCGCCACACAGGGGAGCTTCGCCACGSTGTC------CTC 350

Qy        286 CGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGT 345
                  |      |||      |||      |||  |  |  |     || | | |  |     |
Db        351 TTCCTTCCCGTCCCTGTCGGCTGCCAACCTGCTCATCATCACCGGCGCCTTTGTCATGGC 410

Qy        346 GCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTT 405
                | ||||| |  |||||| | || ||  ||  |||||||  | ||| ||     |||
Db        411 CATCGGCTTCGTGGGCTGCCTGGGTGCCATCAAGGAGAACAAGTGCCTCCTGCTCACTTT 470

Qy        406 CTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGT 465
              ||   || | ||    ||  | ||| | ||||||    | |  |  |||  |   |||||
Db        471 CTTCCTGCTGCTGCTGCTGGTGTTCCTGCTGGAGGCCACCATCGCCATCCTCTTCTTCGC 530

Qy        466 ATTCAAGGACTGGATT 481
                | || ||||  ||||
Db        531 CTACACGGACAAGATT 546


RESULT 10
US-09-333-599-5
; Sequence 5, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
;  APPLICANT: Testa, Jacqueline E.
;  APPLICANT: Quigley, James P.
;  APPLICANT: Seandel, Marco
;  TITLE OF INVENTION: MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE: SUNY
;  CURRENT APPLICATION NUMBER: US/09/333,599
;  CURRENT FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 827
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (57)..(815)
US-09-333-599-5

  Query Match             2.3%;  Score 57.2;  DB 3;  Length 827;
  Best Local Similarity   48.3%;  Pred. No. 2.5e-05;
  Matches  195;  Conservative    0;  Mismatches  203;  Indels    6;  Gaps    1;

Qy        115 CACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATA 174
              | ||| |  |||  ||  | |  |      ||  | |    | |  |   ||| || ||
Db         50 CCCCAGGATGGGTGAGTTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTA 109

Qy        175 CTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGG 234
              | | ||||||  || |||       ||||| ||   ||          || ||||  | ||
Db        110 CCTGCTGTTTACCTACAATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGG 169

Qy        235 CCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGG 294
              | ||||| |   ||     ||| |||       | |||||     |||| ||
Db        170 CATCTGGACGCTGGCCCTCAAGAGTG------ACTACATCAGCCTGCTGGCCTCAGGCAC 223

Qy        295 AGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTT 354
              |||  |  |  |   |  | || ||||  ||    |||  |    |||   ||  |
Db        224 CTACCTGGCCACAGCCTACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGT 283

Qy        355 TGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTT 414
              ||||||  ||   |  ||  ||||          |||||     | |  ||   |  |
Db        284 CTTGGGCTGCTGCGCCACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCT 343

Qy        415 CCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGA 474
              ||   | |||||||  | ||||||      | || ||| | |||| ||  | | |  |
Db        344 GCTCCTCATCATCTTTCTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTACTACCA 403

Qy        475 CTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCA 518
                | |        |||||||      | | ||  ||| | | |
Db        404 GCAGCTGAACACGGAGCTCAAGGAGAACCTGAAGGACACCATGA 447


RESULT 11
US-09-499-781-5
; Sequence 5, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
;  APPLICANT:  Testa, Jacqueline E.
;  APPLICANT:  Quigley, James P.
;  APPLICANT:  Seandel, Marco
;  TITLE OF INVENTION:  MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION:  ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE:  SUNY
;  CURRENT APPLICATION NUMBER: US/09/499,781
;  CURRENT FILING DATE:  2000-02-08
;  PRIOR APPLICATION NUMBER:  09/333,599
;  PRIOR FILING DATE:  1999-06-15
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 827
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (57)..(815)
US-09-499-781-5

  Query Match             2.3%;  Score 57.2;  DB 4;  Length 827;
  Best Local Similarity   48.3%;  Pred. No. 2.5e-05;
  Matches  195;  Conservative    0;  Mismatches  203;  Indels    6;  Gaps    1;

Qy        115 CACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATA 174
              | ||| |  |||  ||  | |  |      ||  | |    | |  |   ||| || ||
Db         50 CCCCAGGATGGGTGAGTTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTA 109

Qy        175 CTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGG 234
              | | ||||||  || |||       ||||| ||   ||          || ||||  | ||
Db        110 CCTGCTGTTTACCTACAATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGG 169

Qy        235 CCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGG 294
              | ||||| |   ||     ||| |||       | |||||     |||| ||
Db        170 CATCTGGACGCTGGCCCTCAAGAGTG------ACTACATCAGCCTGCTGGCCTCAGGCAC 223

Qy        295 AGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTT 354
              |||  |  |  |   |  | || ||||  ||    |||  |    |||   ||  |
Db        224 CTACCTGGCCACAGCCTACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGT 283

Qy        355 TGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTT 414
```

```
Db        284 CTTGGGCTGCTGCGCCACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCT 343

Qy        415 CCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGA 474

Db        344 GCTCCTCATCATCTTTCTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTACTACCA 403

Qy        475 CTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCA 518

Db        404 GCAGCTGAACACGGAGCTCAAGGAGAACCTGAAGGACACCATGA 447


RESULT 12
US-09-333-599-1
; Sequence 1, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
;  APPLICANT: Testa, Jacqueline E.
;  APPLICANT: Quigley, James P.
;  APPLICANT: Seandel, Marco
;  TITLE OF INVENTION: MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE: SUNY
;  CURRENT APPLICATION NUMBER: US/09/333,599
;  CURRENT FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-333-599-1

  Query Match             2.3%;  Score 57.2;  DB 3;  Length 870;
  Best Local Similarity   48.3%;  Pred. No. 2.6e-05;
  Matches  195;  Conservative    0;  Mismatches  203;  Indels    6;  Gaps    1;

Qy        115 CACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATA 174

Db         94 CCCCAGGATGGGTGAGTTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTA 153

Qy        175 CTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGG 234

Db        154 CCTGCTGTTTACCTACAATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGG 213

Qy        235 CCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGG 294

Db        214 CATCTGGACGCTGGCCCTCAAGAGTG------ACTACATCAGCCTGCTGGCCTCAGGCAC 267

Qy        295 AGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTT 354

Db        268 CTACCTGGCCACAGCCTACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGT 327

Qy        355 TGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTT 414

Db        328 CTTGGGCTGCTGCGCCACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCT 387

Qy        415 CCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGA 474

Db        388 GCTCCTCATCATCTTTCTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTACTACCA 447

Qy        475 CTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCA 518

Db        448 GCAGCTGAACACGGAGCTCAAGGAGAACCTGAAGGACACCATGA 491


RESULT 13
US-09-499-781-1
; Sequence 1, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
;  APPLICANT:  Testa, Jacqueline E.
;  APPLICANT:  Quigley, James P.
;  APPLICANT:  Seandel, Marco
;  TITLE OF INVENTION:  MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION:  ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE:  SUNY
;  CURRENT APPLICATION NUMBER: US/09/499,781
;  CURRENT FILING DATE:  2000-02-08
;  PRIOR APPLICATION NUMBER:  09/333,599
;  PRIOR FILING DATE:  1999-06-15
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-499-781-1

  Query Match             2.3%;  Score 57.2;  DB 4;  Length 870;
  Best Local Similarity   48.3%;  Pred. No. 2.6e-05;
  Matches  195;  Conservative    0;  Mismatches  203;  Indels    6;  Gaps    1;

Qy        115 CACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATA 174

Db         94 CCCCAGGATGGGTGAGTTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTA 153

Qy        175 CTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGG 234

Db        154 CCTGCTGTTTACCTACAATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGG 213

Qy        235 CCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGG 294

Db        214 CATCTGGACGCTGGCCCTCAAGAGTG------ACTACATCAGCCTGCTGGCCTCAGGCAC 267

Qy        295 AGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTT 354

Db        268 CTACCTGGCCACAGCCTACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGT 327

Qy        355 TGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTT 414

Db        328 CTTGGGCTGCTGCGCCACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCT 387

Qy        415 CCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGA 474

Db        388 GCTCCTCATCATCTTTCTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTACTACCA 447

Qy        475 CTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCA 518

Db        448 GCAGCTGAACACGGAGCTCAAGGAGAACCTGAAGGACACCATGA 491


RESULT 14
US-08-705-771-8
; Sequence 8, Application US/08705771
; Patent No. 6054289
;  GENERAL INFORMATION:
;    APPLICANT:  Paul Moore, Reiner Gentz, Hongjin Ji,
;    APPLICANT:  Jian Ni and Jing-Shan Hu
;    TITLE OF INVENTION:  Human Genes, Sequences and
;    TITLE OF INVENTION:  Expression Products
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/705,771
;       FILING DATE:  August 30, 1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MULLINS, J.G.
;       REGISTRATION NUMBER:  33,073
;       REFERENCE/DOCKET NUMBER:  325800-346 (PF196)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  973-994-1700
;       TELEFAX:  973-994-1744
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1344 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-705-771-8

  Query Match             2.3%;  Score 57.2;  DB 3;  Length 1344;
  Best Local Similarity   51.3%;  Pred. No. 3.2e-05;
  Matches  162;  Conservative    0;  Mismatches  148;  Indels    6;  Gaps    1;

Qy        166 CGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCT 225
              ||  || ||| || |||| | ||||||| |  | |||||| ||   ||        | ||
Db        167 CGTCAAGTACCTCATGTTCGCCTTCAACCTGCTCTTCTGGCTGGGAGGCTGTGGCGTGCT 226

Qy        226 GGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGAC 285
              ||   ||||| |||||       |   | ||||     | |||   | ||        |
Db        227 GGGTGTCGGCATCTGGCTGGCCGCACAACAGGGGAGCTTTGCCACGCTGTC------CTC 280

Qy        286 CGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGT 345
                  |     |||     |||     ||| |  |  |    || | | |  |     |
Db        281 TTCCTTCCCGTCCCTGTGGGCTGCCAACCTGCTCATCATCACCGGCGCCTTTGTCATGGC 340

Qy        346 GCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTT 405
                | ||||| |  |||||| | || ||  ||  |||||||| | ||| ||      |||
Db        341 CATCGGCTTCGTGGGCTGCCTGGGTGCCATCAAGGAGAACAAGTGCCTCCTGCTCACTTT 400

Qy        406 CTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGT 465
              ||   || | ||    ||  | ||| | ||||||    | |  |  ||| |   |||||
Db        401 CTTCCTGCTGCTGCTGCTGGTGTTCCTGCTGGAGGGCACCATCGCCATCCTCTTCTTCGC 460

Qy        466 ATTCAAGGACTGGATT 481
                | || ||||  ||||
Db        461 CTACACGGACAAGATT 476


RESULT 15
US-08-855-140-2
; Sequence 2, Application US/08855140
; Patent No. 5854022
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Guegler, Karl J.
;    TITLE OF INVENTION:  NEW TRANSMEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,140
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0296 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  MYOMNOT01
;      CLONE:  779308
US-08-855-140-2

  Query Match             2.2%;  Score 54.8;  DB 2;  Length 977;
  Best Local Similarity   50.0%;  Pred. No. 0.00012;
  Matches  168;  Conservative    0;  Mismatches  162;  Indels    6;  Gaps    1;

Qy        160 CTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCT 219
              |||||||  ||| ||| |  ||||   ||||||  |  | ||||||  |     ||
Db        139 CTGCTGCTTGAAGTACATGATGTTCCTCTTCAATTTGATATTCTGGCTCTGTGGCTGTGG 198

Qy        220 GTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGC 279
              | | ||||   | ||| |||||  ||  |       | ||      |  ||| | ||||  |
Db        199 GCTGCTGGGAGTGGGCATCTGGCTCTCCGTGTCCCAAGGCAACTTTGCCACCTTCTCCCC 258

Qy        280 GCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCAT 339
                  |  |  |       |  | |     ||| |  | |   | ||   | |  |
Db        259 CAGCTTCCCTTCG------TTGTCTGCAGCCAACCTGGTCATCGCCATAGGCACCATTGT 312

Qy        340 GTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAA 399
                |||  ||||||   ||||||| | |||||  ||  ||| |||| | ||| |||
Db        313 CATGGTGACGGGCTTCCTCGGCTGCCTGGGGGCCATCAAGGAAAACAAGTGCCTCCTCCT 372

Qy        400 GTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGC 459
                || |   |  ||||      ||| | || |    ||||||      |   ||| |
Db        373 CAGCTTTTTCATCGTCCTGTTGGTCACCCTCCTAGCAGAGCTGATCTTACCCATCCTCTT 432

Qy        460 CTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAA 495
              ||| || | || ||||  | |    ||  |   |||
Db        433 CTTTGTCTACATGGACAAGGTGAACGAGAACGCCAA 468


Search completed: November 21, 2003, 21:33:19
Job time : 152 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:43:38 ; Search time 30 Seconds
(without alignments)
1417.879 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1283	100.0	233	10	US-09-972-970-4	Sequence 4, Appli
2	1066	83.1	270	10	US-09-934-268-2	Sequence 2, Appli
3	1066	83.1	270	15	US-10-103-196-16	Sequence 16, Appl
4	1066	83.1	270	15	US-10-162-435-36	Sequence 36, Appl
5	885	69.0	271	9	US-09-925-299-978	Sequence 978, App
6	885	69.0	271	11	US-09-925-299-978	Sequence 978, App
7	791	61.7	268	15	US-10-103-196-17	Sequence 17, Appl
8	699	54.5	193	9	US-09-729-674-170	Sequence 170, App
9	638.5	49.8	270	11	US-09-796-753-126	Sequence 126, App
10	638.5	49.8	270	11	US-09-905-674-2	Sequence 2, Appli
11	638.5	49.8	270	11	US-09-892-877-267	Sequence 267, App
12	638.5	49.8	270	11	US-09-948-783-269	Sequence 269, App
13	625.5	48.8	270	11	US-09-892-877-126	Sequence 126, App
14	625.5	48.8	270	11	US-09-948-783-127	Sequence 127, App
15	619.5	48.3	270	11	US-09-957-187-16	Sequence 16, Appl
16	482.5	37.6	122	14	US-10-042-417-38	Sequence 38, Appl
17	429	33.4	120	15	US-10-106-698-6296	Sequence 6296, Ap
18	412.5	32.2	245	11	US-09-984-271-133	Sequence 133, App
19	411.5	32.1	209	11	US-09-957-187-79	Sequence 79, Appl
20	383	29.9	111	15	US-10-103-196-22	Sequence 22, Appl
21	325	25.3	254	10	US-09-934-268-4	Sequence 4, Appli
22	325	25.3	254	15	US-10-162-435-38	Sequence 38, Appl
23	324	25.3	294	11	US-09-946-374-123	Sequence 123, App
24	324	25.3	294	12	US-10-015-387A-123	Sequence 123, App
25	324	25.3	294	12	US-10-006-130A-123	Sequence 123, App
26	324	25.3	294	12	US-10-199-672-278	Sequence 278, App
27	324	25.3	294	12	US-10-006-172A-123	Sequence 123, App
28	324	25.3	294	12	US-10-187-749-278	Sequence 278, App
29	324	25.3	294	12	US-10-194-457-278	Sequence 278, App
30	324	25.3	294	12	US-10-184-642-278	Sequence 278, App
31	324	25.3	294	12	US-10-196-747-278	Sequence 278, App
32	324	25.3	294	12	US-10-015-392A-123	Sequence 123, App
33	324	25.3	294	12	US-10-017-253A-123	Sequence 123, App
34	324	25.3	294	12	US-10-173-689-278	Sequence 278, App
35	324	25.3	294	12	US-10-173-690-278	Sequence 278, App
36	324	25.3	294	12	US-10-173-691-278	Sequence 278, App
37	324	25.3	294	12	US-10-173-692-278	Sequence 278, App
38	324	25.3	294	12	US-10-173-694-278	Sequence 278, App
39	324	25.3	294	12	US-10-173-698-278	Sequence 278, App
40	324	25.3	294	12	US-10-173-699-278	Sequence 278, App
41	324	25.3	294	12	US-10-173-707-278	Sequence 278, App
42	324	25.3	294	12	US-10-174-569-278	Sequence 278, App
43	324	25.3	294	12	US-10-174-583-278	Sequence 278, App
44	324	25.3	294	12	US-10-174-587-278	Sequence 278, App
45	324	25.3	294	12	US-10-174-589-278	Sequence 278, App

ALIGNMENTS

```
RESULT 1
US-09-972-970-4
; Sequence 4, Application US/09972970
; Patent No. US20020164693A1
; GENERAL INFORMATION:
;  APPLICANT: Shi et al.
;  TITLE OF INVENTION: TM4SF Receptor Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE: PT056P1
;  CURRENT APPLICATION NUMBER: US/09/972,970
;  CURRENT FILING DATE:  2001-10-10
;  PRIOR APPLICATION NUMBER: PCT/US01/11130
;  PRIOR FILING DATE: 2001-04-05
;  PRIOR APPLICATION NUMBER: 60/195,336
;  PRIOR FILING DATE: 2000-04-10
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 233
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-972-970-4

  Query Match           100.0%;  Score 1283;  DB 10;  Length 233;
  Best Local Similarity  100.0%;  Pred. No. 9.7e-119;
  Matches  233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
```

```
Qy        181 RCGVPFSCCVRDPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGCRTT 233
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RCGVPFSCCVRDPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGCRTT 233


RESULT 2
US-09-934-268-2
; Sequence 2, Application US/09934268
; Patent No. US20020172986A1
; GENERAL INFORMATION:
;  APPLICANT: Leiby, Kevin R.
;  TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY
;  TITLE OF INVENTION:  MEMBER AND USES THEREOF
;  FILE REFERENCE: 10448-079001
;  CURRENT APPLICATION NUMBER: US/09/934,268
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-934-268-2

  Query Match             83.1%;  Score 1066;  DB 10;  Length 270;
  Best Local Similarity   99.5%;  Pred. No. 3.2e-97;
  Matches  193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 3
US-10-103-196-16
; Sequence 16, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE: PT008P1
;  CURRENT APPLICATION NUMBER: US/10/103,196
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/707,936
;  PRIOR FILING DATE: 2000-11-08
;  PRIOR APPLICATION NUMBER: PCT/US00/13504
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: 60/178,770
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/149,447
;  PRIOR FILING DATE: 1999-08-18
;  PRIOR APPLICATION NUMBER: 60/138,573
;  PRIOR FILING DATE: 1999-06-11
;  PRIOR APPLICATION NUMBER: 60/137,797
;  PRIOR FILING DATE: 1999-06-03
;  PRIOR APPLICATION NUMBER: 60/135,122
;  PRIOR FILING DATE: 1999-05-19
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-103-196-16

  Query Match             83.1%;  Score 1066;  DB 15;  Length 270;
  Best Local Similarity   99.5%;  Pred. No. 3.2e-97;
  Matches  193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 4
US-10-162-435-36
; Sequence 36, Application US/10162435
; Publication No. US20030096305A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Curtis , Rory A. J.
;  APPLICANT:  Kapeller-Libermann, Rosana
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Leiby , Kevin R.
;  TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND
;  TITLE OF INVENTION:  CELL SURFACE PROTEIN FAMILY MEMBERS
;  FILE REFERENCE: 10448-189001
;  CURRENT APPLICATION NUMBER: US/10/162,435
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US 09/836,499
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: PCT/US01/12420
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: US 60/197,507
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: US 09/891,008
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: PCT/US01/19963
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/214,220
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 09/860,868
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: PCT/US01/16013
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,674
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/886,429
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: PCT/US01/20055
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/213,963
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 10/041,406
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: PCT/US02/00275
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/260,286
;  PRIOR FILING DATE: 2001-01-08
```

```
;  PRIOR APPLICATION NUMBER: US 09/934,268
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/41811
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-162-435-36

  Query Match             83.1%;  Score 1066;  DB 15;  Length 270;
  Best Local Similarity   99.5%;  Pred. No. 3.2e-97;
  Matches  193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 5
US-09-925-299-978
; Sequence 978, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 978
;   LENGTH: 271
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-299-978

  Query Match             69.0%;  Score 885;  DB 9;  Length 271;
  Best Local Similarity   79.9%;  Pred. No. 2.5e-79;
  Matches  155;  Conservative   16;  Mismatches   21;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| ||||||||::|||||
Db          4 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 61

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         62 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 121

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::|||||||||||||| |||| |||| | :||||||||||| | |||
Db        122 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 181

Qy        181 RCGVPFSCCVRDPA 194
              |||||||||  :|||
Db        182 RCGVPFSCCTKDPA 195


RESULT 6
US-09-925-299-978
; Sequence 978, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 978
;   LENGTH: 271
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-299-978

  Query Match             69.0%;  Score 885;  DB 11;  Length 271;
  Best Local Similarity   79.9%;  Pred. No. 2.5e-79;
  Matches  155;  Conservative   16;  Mismatches   21;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| ||||||||::|||||
Db          4 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 61

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         62 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 121

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::|||||||||||||| |||| |||| | :||||||||||| | |||
Db        122 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 181

Qy        181 RCGVPFSCCVRDPA 194
              |||||||||  :|||
Db        182 RCGVPFSCCTKDPA 195


RESULT 7
US-10-103-196-17
; Sequence 17, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE: PT008P1
;  CURRENT APPLICATION NUMBER: US/10/103,196
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/707,936
;  PRIOR FILING DATE: 2000-11-08
;  PRIOR APPLICATION NUMBER: PCT/US00/13504
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: 60/178,770
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/149,447
;  PRIOR FILING DATE: 1999-08-18
;  PRIOR APPLICATION NUMBER: 60/138,573
;  PRIOR FILING DATE: 1999-06-11
;  PRIOR APPLICATION NUMBER: 60/137,797
;  PRIOR FILING DATE: 1999-06-03
;  PRIOR APPLICATION NUMBER: 60/135,122
;  PRIOR FILING DATE: 1999-05-19
;  NUMBER OF SEQ ID NOS: 44
```

```
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 268
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-103-196-17

  Query Match             61.7%;  Score 791;  DB 15;  Length 268;
  Best Local Similarity   72.2%;  Pred. No. 4.7e-70;
  Matches  140;  Conservative   16;  Mismatches   36;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| |||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
               ||||||:|||||| :|||||||||  :      ||      |  ||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGAYGKTLLSFSSFFLCSWELFSFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::|||||||||||||| |||| |||| | :|||||||||||| | |||
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 178

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||| :|||
Db        179 RCGVPFSCCTKDPA 192


RESULT 8
US-09-729-674-170
; Sequence 170, Application US/09729674
; Patent No. US20010039335A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Steininger II, Robert J.
;  APPLICANT:  Spaulding, Vikki
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6055-64X
;  CURRENT APPLICATION NUMBER: US/09/729,674
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: 09/539,330
;  PRIOR FILING DATE: 2000-03-30
;  NUMBER OF SEQ ID NOS: 283
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
;   LENGTH: 193
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-729-674-170

  Query Match             54.5%;  Score 699;  DB 9;  Length 193;
  Best Local Similarity   62.9%;  Pred. No. 3.9e-61;
  Matches  139;  Conservative   16;  Mismatches   30;  Indels   36;  Gaps    4;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| |||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
               ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::|||||||||||||| |||                         | :
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEY-----------------------IPMQV 155

Qy        181 RCGVPFSCCVRDPA-------MSSTPSVAMMSGSNWSWSSR 214
              | |       ||       |||| ||||| | |   :||
Db        156 ESDVAF----HSPAALKIPQKMSSTLSVAMMPGKNQKLTSR 192


RESULT 9
US-09-796-753-126
; Sequence 126, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
;  APPLICANT:  McCarthy, Sean A.
;  TITLE OF INVENTION:  SECRETED PROTEINS AND USES THEREOF
;  FILE REFERENCE:  7853-227-999
;  CURRENT APPLICATION NUMBER: US/09/796,753
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER:  09/183,175
;  PRIOR FILING DATE:  1998-10-30
;  PRIOR APPLICATION NUMBER:  09/223,094
;  PRIOR FILING DATE:  1998-12-30
;  PRIOR APPLICATION NUMBER:  09/223,546
;  PRIOR FILING DATE:  1998-12-30
;  PRIOR APPLICATION NUMBER:  09/224,246
;  PRIOR FILING DATE:  1998-12-30
;  PRIOR APPLICATION NUMBER:  09/259,388
;  PRIOR FILING DATE:  1999-02-26
;  PRIOR APPLICATION NUMBER:  60/122,458
;  PRIOR FILING DATE:  1999-03-01
;  PRIOR APPLICATION NUMBER:  09/312,359
;  PRIOR FILING DATE:  1999-05-14
;  PRIOR APPLICATION NUMBER:  09/336,536
;  PRIOR FILING DATE:  1999-06-18
;  PRIOR APPLICATION NUMBER:  09/342,687
;  PRIOR FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER:  09/345,464
;  PRIOR FILING DATE:  1999-06-30
;  PRIOR APPLICATION NUMBER:  09/365,164
;  PRIOR FILING DATE:  1999-07-30
;  PRIOR APPLICATION NUMBER:  09/399,723
;  PRIOR FILING DATE:  1999-09-20
;  PRIOR APPLICATION NUMBER:  09/409,634
;  PRIOR FILING DATE:  1999-09-30
;  PRIOR APPLICATION NUMBER:  09/471,179
;  PRIOR FILING DATE:  1999-12-23
;  PRIOR APPLICATION NUMBER:  09/474,071
;  PRIOR FILING DATE:  1999-12-29
;  PRIOR APPLICATION NUMBER:  09/474,072
;  PRIOR FILING DATE:  1999-12-29
;  PRIOR APPLICATION NUMBER:  09/514,010
;  PRIOR FILING DATE:  2000-02-25
;  PRIOR APPLICATION NUMBER:  09/516,745
;  PRIOR FILING DATE:  2000-03-01
;  PRIOR APPLICATION NUMBER:  09/572,002
;  PRIOR FILING DATE:  2000-05-14
;  PRIOR APPLICATION NUMBER:  09/597,993
;  PRIOR FILING DATE:  2000-06-19
;  PRIOR APPLICATION NUMBER:  09/599,596
;  PRIOR FILING DATE:  2000-06-22
;  PRIOR APPLICATION NUMBER:  09/630,334
;  PRIOR FILING DATE:  2000-07-31
;  PRIOR APPLICATION NUMBER:  09/606,565
;  PRIOR FILING DATE:  2000-06-29
;  PRIOR APPLICATION NUMBER:  09/606,317
;  PRIOR FILING DATE:  2000-06-29
;  PRIOR APPLICATION NUMBER:  09/665,666
;  PRIOR FILING DATE:  2000-09-20
;  PRIOR APPLICATION NUMBER:  09/677,751
;  PRIOR FILING DATE:  2000-09-30
;  NUMBER OF SEQ ID NOS:  162
; SEQ ID NO 126
```

```
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-796-753-126

  Query Match             49.8%;  Score 638.5;  DB 11;  Length 270;
  Best Local Similarity   49.4%;  Pred. No. 5.5e-55;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :|||||| |||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  |||||:|||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
                 |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :   | |              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 10
US-09-905-674-2
; Sequence 2, Application US/09905674
; Publication No. US20030039647A1
; GENERAL INFORMATION:
;  APPLICANT: Reinhard, Christoph
;  APPLICANT:  Garcia, Pablo
;  TITLE OF INVENTION: TETRASPAN PROTEIN AND USES THEREOF
;  FILE REFERENCE: PP-01700.002/200130.521
;  CURRENT APPLICATION NUMBER: US/09/905,674
;  CURRENT FILING DATE:  2001-07-13
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-905-674-2

  Query Match             49.8%;  Score 638.5;  DB 11;  Length 270;
  Best Local Similarity   49.4%;  Pred. No. 5.5e-55;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :|||||| |||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  |||||:|||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
                 |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :   | |              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 11
US-09-892-877-267
; Sequence 267, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et. al.
;  TITLE OF INVENTION: 97 Human secreted proteins
;  FILE REFERENCE: PZ028P1
;  CURRENT APPLICATION NUMBER: US/09/892,877
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
;  NUMBER OF SEQ ID NOS: 461
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 267
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-892-877-267

  Query Match             49.8%;  Score 638.5;  DB 11;  Length 270;
  Best Local Similarity   49.4%;  Pred. No. 5.5e-55;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :|||||| |||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  |||||:|||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
                 |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :   | |              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 12
US-09-948-783-269
; Sequence 269, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et. al.
;  TITLE OF INVENTION: 97 Human secreted proteins
;  FILE REFERENCE: PZ028P2
;  CURRENT APPLICATION NUMBER: US/09/948,783
;  CURRENT FILING DATE:  2001-09-10
;  PRIOR APPLICATION NUMBER: 60/231,846
;  PRIOR FILING DATE: 2000-09-11
;  PRIOR APPLICATION NUMBER: 09/892,877
;  PRIOR FILING DATE: 2001-06-28
;  PRIOR APPLICATION NUMBER: 09/437,658
;  PRIOR FILING DATE: 1999-11-10
;  PRIOR APPLICATION NUMBER: PCT/US99/09847
;  PRIOR FILING DATE: 1999-05-06
;  PRIOR APPLICATION NUMBER: 60/085,093
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,094
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,105
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,180
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,927
```

```
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,906
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,924
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,922
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,921
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,923
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,925
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,928
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,920
;  PRIOR FILING DATE: 1998-05-18
;  NUMBER OF SEQ ID NOS: 465
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 269
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-948-783-269

  Query Match             49.8%;  Score 638.5;  DB 11;  Length 270;
  Best Local Similarity   49.4%;  Pred. No. 5.5e-55;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :||||| ||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:||||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
              |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :    | :   | |              || |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 13
US-09-892-877-126
; Sequence 126, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et. al.
;  TITLE OF INVENTION: 97 Human secreted proteins
;  FILE REFERENCE: PZ028P1
;  CURRENT APPLICATION NUMBER: US/09/892,877
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
;  NUMBER OF SEQ ID NOS: 461
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 126
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (27)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-892-877-126

  Query Match             48.8%;  Score 625.5;  DB 11;  Length 270;
  Best Local Similarity   49.0%;  Pred. No. 1.1e-53;
  Matches  118;  Conservative   37;  Mismatches   67;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:| : | :|| :||||| |||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFXLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:||||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
              |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :    | :   | |              || |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 14
US-09-948-783-127
; Sequence 127, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et. al.
;  TITLE OF INVENTION: 97 Human secreted proteins
;  FILE REFERENCE: PZ028P2
;  CURRENT APPLICATION NUMBER: US/09/948,783
;  CURRENT FILING DATE:  2001-09-10
;  PRIOR APPLICATION NUMBER: 60/231,846
;  PRIOR FILING DATE: 2000-09-11
;  PRIOR APPLICATION NUMBER: 09/892,877
;  PRIOR FILING DATE: 2001-06-28
;  PRIOR APPLICATION NUMBER: 09/437,658
;  PRIOR FILING DATE: 1999-11-10
;  PRIOR APPLICATION NUMBER: PCT/US99/09847
;  PRIOR FILING DATE: 1999-05-06
;  PRIOR APPLICATION NUMBER: 60/085,093
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,094
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,105
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,180
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,927
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,906
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,924
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,922
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,921
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,923
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,925
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,928
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,920
;  PRIOR FILING DATE: 1998-05-18
;  NUMBER OF SEQ ID NOS: 465
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 127
```

```
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (27)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-948-783-127

  Query Match             48.8%;  Score 625.5;  DB 11;  Length 270;
  Best Local Similarity   49.0%;  Pred. No. 1.1e-53;
  Matches  118;  Conservative   37;  Mismatches   67;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:| : | :|| :||||| |||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFXLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:||||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
              |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW----------SWSSRAPYTPKAVWAS 225
              |||||| |||      :     | :   | |              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 15
US-09-957-187-16
; Sequence 16, Application US/09957187
; Publication No. US20030054514A1
; GENERAL INFORMATION:
;  APPLICANT: Shimkets, Richard A.
;  APPLICANT:  LaRochelle, William
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND PROTEINS ENCODED THEREBY
;  FILE REFERENCE: 15966-540 CIP
;  CURRENT APPLICATION NUMBER: US/09/957,187
;  CURRENT FILING DATE:  2000-09-19
;  PRIOR APPLICATION NUMBER: 60/123,667
;  PRIOR FILING DATE: 1999-03-09
;  PRIOR APPLICATION NUMBER: 09/520,781
;  PRIOR FILING DATE: 2000-03-03
;  PRIOR APPLICATION NUMBER: 60/234,082
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: 60/233,798
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: 60/174,485
;  PRIOR FILING DATE: 2000-01-04
;  NUMBER OF SEQ ID NOS: 85
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 270
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-957-187-16

  Query Match             48.3%;  Score 619.5;  DB 11;  Length 270;
  Best Local Similarity   48.1%;  Pred. No. 4.2e-53;
  Matches  116;  Conservative   37;  Mismatches   69;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :||||| |||||::: :| : |:||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPA 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||| :|| |||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGAVMFTLGFAGRVGARRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
              |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW----------SWSSRAPYTPKAVWAS 225
              |||||| |||      :     | :   | |              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


Search completed: November 21, 2003, 13:46:43
Job time : 31 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 17:20:09 ; Search time 770 Seconds
(without alignments)
10772.341 Million cell updates/sec

Title: US-09-972-970-2
Perfect score: 2538
Sequence: 1 ccacgcgtccggccgcagcc.....aaaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

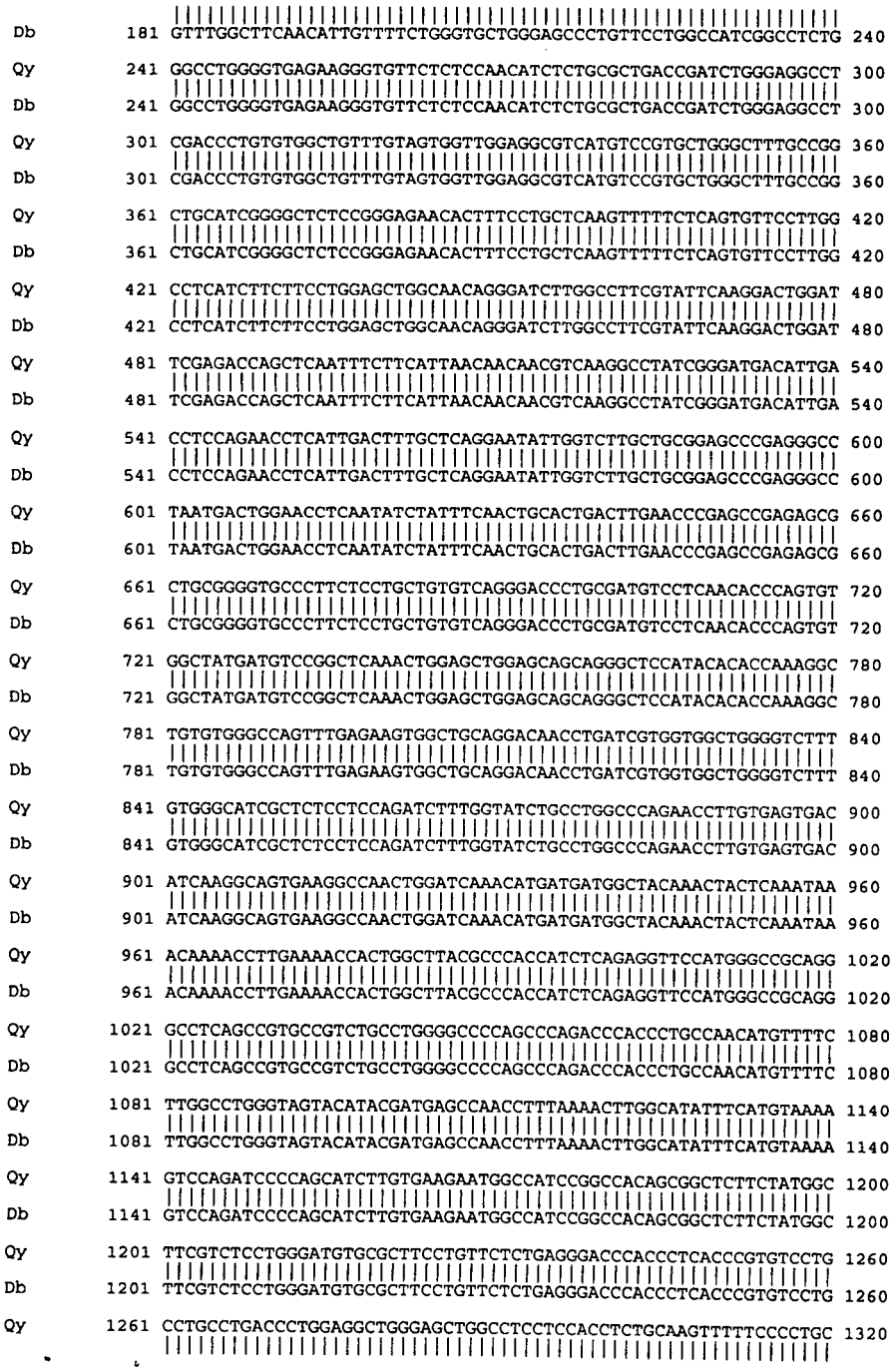
Result No.	Score	% Query Match	Length	DB	ID	Description
1	2538	100.0	2538	10	US-09-972-970-2	Sequence 2, Appli
2	1061.6	41.8	2461	12	US-10-252-157-153	Sequence 153, App
3	790.4	31.1	3184	10	US-09-934-268-1	Sequence 1, Appli
4	790.4	31.1	3184	14	US-10-162-435-35	Sequence 35, Appl
5	787.2	31.0	1126	14	US-10-103-196-4	Sequence 4, Appli
6	710.4	28.0	813	10	US-09-934-268-3	Sequence 3, Appli
7	710.4	28.0	813	14	US-10-162-435-37	Sequence 37, Appl
8	441.6	17.4	1655	9	US-09-925-299-205	Sequence 205, App
9	441.6	17.4	1655	11	US-09-925-299-205	Sequence 205, App
10	418.6	16.5	1174	14	US-10-103-196-5	Sequence 5, Appli
11	402.2	15.8	1178	14	US-10-103-196-10	Sequence 10, Appl
12	390.8	15.4	1988	11	US-09-822-846-126	Sequence 126, App
13	357.4	14.1	2672	11	US-09-892-877-20	Sequence 20, Appl
14	357.4	14.1	2672	11	US-09-948-783-20	Sequence 20, Appl
15	353.4	13.9	2715	11	US-09-796-753-125	Sequence 125, App
16	351.8	13.9	1388	11	US-09-905-674-1	Sequence 1, Appli
17	343.8	13.5	1428	11	US-09-991-053-15	Sequence 15, Appl
18	343.8	13.5	1428	11	US-09-957-187-15	Sequence 15, Appl
19	333.6	13.1	1110	9	US-09-729-674-169	Sequence 169, App
20	321	12.6	1667	11	US-09-892-877-266	Sequence 266, App
21	321	12.6	1667	11	US-09-948-783-268	Sequence 268, App
22	295.8	11.7	864	10	US-09-875-440-1	Sequence 1, Appli
23	294.2	11.6	368	13	US-10-042-417-37	Sequence 37, Appl
24	254.4	10.0	816	11	US-09-991-053-78	Sequence 78, Appl
25	254.4	10.0	816	11	US-09-957-187-78	Sequence 78, Appl
c 26	240	9.5	839	14	US-10-106-698-2019	Sequence 2019, Ap
27	234	9.2	1932	11	US-09-984-271-41	Sequence 41, Appl
28	231.6	9.1	361	11	US-09-918-995-37420	Sequence 37420, A
29	230	9.1	411	11	US-09-918-995-7371	Sequence 7371, Ap
30	200.2	7.9	360	11	US-09-918-995-37648	Sequence 37648, A
31	152.4	6.0	476	11	US-09-918-995-31988	Sequence 31988, A
32	130.8	5.2	476	11	US-09-918-995-36537	Sequence 36537, A
33	128	5.0	1725	14	US-10-103-196-7	Sequence 7, Appli
34	127	5.0	1772	11	US-09-813-153-58	Sequence 58, Appl
35	126.2	5.0	1778	11	US-09-946-374-122	Sequence 122, App
36	126.2	5.0	1778	12	US-10-015-387A-122	Sequence 122, App
37	126.2	5.0	1778	12	US-10-006-130A-122	Sequence 122, App
38	126.2	5.0	1778	12	US-10-199-672-277	Sequence 277, App
39	126.2	5.0	1778	12	US-10-006-172A-122	Sequence 122, App
40	126.2	5.0	1778	12	US-10-187-749-277	Sequence 277, App
41	126.2	5.0	1778	12	US-10-194-457-277	Sequence 277, App
42	126.2	5.0	1778	12	US-10-184-642-277	Sequence 277, App
43	126.2	5.0	1778	12	US-10-196-747-277	Sequence 277, App
44	126.2	5.0	1778	12	US-10-015-392A-122	Sequence 122, App
45	126.2	5.0	1778	12	US-10-017-253A-122	Sequence 122, App

ALIGNMENTS

RESULT 1
US-09-972-970-2
; Sequence 2, Application US/09972970
; Patent No. US20020164693A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: TM4SF Receptor Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT056P1
; CURRENT APPLICATION NUMBER: US/09/972,970
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11130
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/195,336
; PRIOR FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-972-970-2

Query Match 100.0%; Score 2538; DB 10; Length 2538;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CCACGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCACGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCC 60

Qy         61 GTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCAT 120

Qy        121 GCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCT 180

Qy        181 GTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTG 240
```


Db 181 GTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTG 240
Qy 241 GGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCT 300
Db 241 GGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCT 300
Qy 301 CGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGG 360
Db 301 CGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGG 360
Qy 361 CTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGG 420
Db 361 CTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGG 420
Qy 421 CCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGAT 480
Db 421 CCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGAT 480
Qy 481 TCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGA 540
Db 481 TCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGA 540
Qy 541 CCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCC 600
Db 541 CCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCC 600
Qy 601 TAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCG 660
Db 601 TAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCG 660
Qy 661 CTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTCAACACCCAGTGT 720
Db 661 CTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTCAACACCCAGTGT 720
Qy 721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGC 780
Db 721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGC 780
Qy 781 TGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTT 840
Db 781 TGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTT 840
Qy 841 GTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGAC 900
Db 841 GTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGAC 900
Qy 901 ATCAAGGCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAA 960
Db 901 ATCAAGGCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAA 960
Qy 961 ACAAAACCTTGAAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGG 1020
Db 961 ACAAAACCTTGAAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGG 1020
Qy 1021 GCCTCAGCCGTGCCGTCTGCCTGGGGCCCCAGCCCAGACCCACCCTGCCAACATGTTTTC 1080
Db 1021 GCCTCAGCCGTGCCGTCTGCCTGGGGCCCCAGCCCAGACCCACCCTGCCAACATGTTTTC 1080
Qy 1081 TTGGCCTGGGTAGTACATACGATGAGCCAACCTTTAAAACTTGGCATATTTCATGTAAAA 1140
Db 1081 TTGGCCTGGGTAGTACATACGATGAGCCAACCTTTAAAACTTGGCATATTTCATGTAAAA 1140
Qy 1141 GTCCAGATCCCCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGC 1200
Db 1141 GTCCAGATCCCCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGC 1200
Qy 1201 TTCGTCTCCTGGGATGTGCGCTTCCTGTTCTCTGAGGGACCCACCCTCACCCGTGTCCTG 1260
Db 1201 TTCGTCTCCTGGGATGTGCGCTTCCTGTTCTCTGAGGGACCCACCCTCACCCGTGTCCTG 1260
Qy 1261 CCTGCCTGACCCTGGAGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTCCCCTGC 1320
Db 1261 CCTGCCTGACCCTGGAGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTCCCCTGC 1320
Qy 1321 AAATGCTGCAAGGCTGCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAGAATTGG 1380
Db 1321 AAATGCTGCAAGGCTGCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAGAATTGG 1380
Qy 1381 GGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCCTGGGGG 1440
Db 1381 GGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCCTGGGGG 1440
Qy 1441 CCAAGATGGGTGCCACCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATCAGGAAG 1500
Db 1441 CCAAGATGGGTGCCACCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATCAGGAAG 1500
Qy 1501 GTTTTGGGCAGGACGTAGCTGGAAGCCTGAGCTTGTCACCCATGGGGATGGGGAGAGCCC 1560
Db 1501 GTTTTGGGCAGGACGTAGCTGGAAGCCTGAGCTTGTCACCCATGGGGATGGGGAGAGCCC 1560
Qy 1561 TGTTTGAGGGCGGCTGATGGTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTTCC 1620
Db 1561 TGTTTGAGGGCGGCTGATGGTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTTCC 1620
Qy 1621 CAGTGGCCTGTAGAGTTGCCTCCTGACCACTAGAGGGCGCGCCCACACAGCATTACCTGG 1680
Db 1621 CAGTGGCCTGTAGAGTTGCCTCCTGACCACTAGAGGGCGCGCCCACACAGCATTACCTGG 1680
Qy 1681 GTCTGCCTTTCCTAGGACAACCCCACCCAGTACAGCCCTGTGCCTGGTGTGTCCACCCTG 1740
Db 1681 GTCTGCCTTTCCTAGGACAACCCCACCCAGTACAGCCCTGTGCCTGGTGTGTCCACCCTG 1740
Qy 1741 CTTACTAGTTCTTTGGGTTTCATGGAATTTACAAGCTTCTAAAGGAGCAGAGTGGCTCAG 1800
Db 1741 CTTACTAGTTCTTTGGGTTTCATGGAATTTACAAGCTTCTAAAGGAGCAGAGTGGCTCAG 1800
Qy 1801 ATTGGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGTGGAGTATTTG 1860
Db 1801 ATTGGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGTGGAGTATTTG 1860
Qy 1861 TGAATCAAAGGAGAGGTTTGGCCTAGTGCCCAGTCTTTTAACTTAGATGCCCTCAGGGCC 1920
Db 1861 TGAATCAAAGGAGAGGTTTGGCCTAGTGCCCAGTCTTTTAACTTAGATGCCCTCAGGGCC 1920
Qy 1921 GGGTGGGTTATAAAAATAAAGTAGGCCTTTGAGCTGTGAGGCCTTTGGGACTTTAATTTT 1980
Db 1921 GGGTGGGTTATAAAAATAAAGTAGGCCTTTGAGCTGTGAGGCCTTTGGGACTTTAATTTT 1980
Qy 1981 TCCCACTATTCCTGGAGATGGGACATAGAGAGACATTGCTTTGTGCTGAGAAATACTTGC 2040
Db 1981 TCCCACTATTCCTGGAGATGGGACATAGAGAGACATTGCTTTGTGCTGAGAAATACTTGC 2040
Qy 2041 ATGATTGAGTCTGAGTCGCTAAGGGCAACTGGCCTTGAGTGACATCAAGGGGTGGTGGGG 2100
Db 2041 ATGATTGAGTCTGAGTCGCTAAGGGCAACTGGCCTTGAGTGACATCAAGGGGTGGTGGGG 2100
Qy 2101 ACTGTGGCAAACCACAGATTCCCACCTGAAATTGGTGGCTGTCCTTCCGTTGGGGCTAAT 2160
Db 2101 ACTGTGGCAAACCACAGATTCCCACCTGAAATTGGTGGCTGTCCTTCCGTTGGGGCTAAT 2160
Qy 2161 GGCTGTACAGCGAGAATGTAGGTAGGTCTGTCTAATGGGAGAAGTCTGGAGAAGCCAAGA 2220
Db 2161 GGCTGTACAGCGAGAATGTAGGTAGGTCTGTCTAATGGGAGAAGTCTGGAGAAGCCAAGA 2220
Qy 2221 AGCTAGATTTTTCATGTGAACTATCCCGAGTTTTAAGTTGTTTGCAGCTAATGAGAAAAA 2280
Db 2221 AGCTAGATTTTTCATGTGAACTATCCCGAGTTTTAAGTTGTTTGCAGCTAATGAGAAAAA 2280
Qy 2281 CCTCTTAAACCCTGATAGTCAAAAGGTGTGGGGGCCATCTTTGACACCTCCCCCACCATA 2340
Db 2281 CCTCTTAAACCCTGATAGTCAAAAGGTGTGGGGGCCATCTTTGACACCTCCCCCACCATA 2340
Qy 2341 GGTCCCTCAGGGACAGTGCCCCATGGGAGCCCTGGTGAGTCCACGGTTCAGGAATGCTGG 2400
Db 2341 GGTCCCTCAGGGACAGTGCCCCATGGGAGCCCTGGTGAGTCCACGGTTCAGGAATGCTGG 2400

```
Qy       2401 GAACTGCTGCAGGTGGGCGGGTTGTGGGCCAGCACCCATCGTGGCTCCCAGGTGTGGGCT 2460
Db       2401 GAACTGCTGCAGGTGGGCGGGTTGTGGGCCAGCACCCATCGTGGCTCCCAGGTGTGGGCT 2460

Qy       2461 GGGCTGGGCTATGGGTGGGCTCTACATGCTACAATAAATGGGGCTCATGATAAAAAAAAA 2520
Db       2461 GGGCTGGGCTATGGGTGGGCTCTACATGCTACAATAAATGGGGCTCATGATAAAAAAAAA 2520

Qy       2521 AAAAAAAAAAAAAAAAAA 2538
Db       2521 AAAAAAAAAAAAAAAAAA 2538
```

```
RESULT 2
US-10-252-157-153
; Sequence 153, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
;  APPLICANT: Faris, Mary
;  APPLICANT:  Pearson, Cecelia I.
;  TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
;  FILE REFERENCE: PA-0027-1 US
;  CURRENT APPLICATION NUMBER: US/10/252,157
;  CURRENT FILING DATE:  2002-10-01
;  PRIOR APPLICATION NUMBER: 60/295,048
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 501
;  SOFTWARE: PERL Program
; SEQ ID NO 153
;   LENGTH: 2461
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030190640A1 234056.5
US-10-252-157-153

  Query Match             41.8%;  Score 1061.6;  DB 12;  Length 2461;
  Best Local Similarity   72.8%;  Pred. No. 3.3e-299;
  Matches 1775;  Conservative    0;  Mismatches  524;  Indels  139;  Gaps   26;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
Db          5 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 64

Qy         72 ATCCCGG--CCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAA 129
Db         65 GGCCTGGCTCCCGGCCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAA 124

Qy        130 GCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTT 189
Db        125 GCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTT 184

Qy        190 CAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGG 249
Db        185 CAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGG 244

Qy        250 TGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGT 309
Db        245 TGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGT 304

Qy        310 GTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGG 369
Db        305 GTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGG 364

Qy        370 GGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTT 429
Db        365 GGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTT 424

Qy        430 CTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCA 489
Db        425 CTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCA 484

Qy        490 GCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAA 549
Db        485 GCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAA 544

Qy        550 CCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTG 609
Db        545 CCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTG 604

Qy        610 GAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGT 669
Db        605 GAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGT 664

Qy        670 GCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTA 725
Db        665 GCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTA 724

Qy        726 TGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGT 785
Db        725 CGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGT 784

Qy        786 GGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGG 845
Db        785 GGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGG 844

Qy        846 CATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAA 905
Db        845 CATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAA 904

Qy        906 GGCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAAACAAA 965
Db        905 GGCAGTGAAAGCCAACTGGAGCAAATGGAATGATG------------------------- 939

Qy        966 ACCTTGAAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGCCGCAGGGCCTC 1025
Db        940 ACTTTGAAAACCACTGGCTTACGCCCACCATTTCCGAGGTCCTGTCCACGGCGGGGCCTC 999

Qy       1026 AGCCGTGCCGTCTGCCTGGGGCCCC-AGCCCAGACCCACCCTGCCAACATGTTTTC-TTG 1083
Db       1000 AGCAGAACTCTCTGACTGGGGCCCCTGGCCCGGCCCCACCCAGCCGACATGTTTTCTTTG 1059

Qy       1084 GCCTGGGTAGTACATACGATGAGCCAACCTTTAAAACTTGGCATATTTCATGTAAAAGTC 1143
Db       1060 GCCTGGGTGGTTTATACCCTGAGCCAACCTTTAAAAATTGGTAGATTTCACATAAAAGTC 1119

Qy       1144 CAGATCCCCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGCTTC 1203
Db       1120 CAGATCCACAGCTTCTCTTGAAGAATGACCACCTGGCTACGCCGGCTCTTCGGTGGCAAC 1179

Qy       1204 GTCTCCTGGGATGTGCGCTTCCTGTTCTCTGAGGGACCCACCCTCACCCGTGTCC----- 1258
Db       1180 ACTACCTGGGACACTGCCTCCCCAGTCACCAAGGGCCCCAGCTGGCCCGTTCTACTCACC 1239

Qy       1259 ----TGCCTGCCTGACCCTGGAGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTC 1314
Db       1240 TAAGTGCCGCCTGACCCTTGTACACTAGGAGCTGGCCTC--CCACCTCTGCAGGGTTATT 1297

Qy       1315 CCCTGCAAATGCTGCAAGGCTGCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAG 1374
Db       1298 TCCTGCA----CCTCGAGGCCGCTGCGGGCCAATCTGGAGTGAAACACGGGGACCTGAAG 1353

Qy       1375 AATTGGGGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCC 1434
Db       1354 GA-TGGAGAGGCTGGACCCCGCTTTGAAGAGGGTGCAGCCTGGGAAGGG--CGGCCTTGC 1410

Qy       1435 TGGGGGCCAAGATGGGTGCCACCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATC 1494
Db       1411 TGGGGACTGCGGTGGGAG-TAGAGTGCCCAGGAGAGGGTCTGAGGGGTGGGATGGGGTC 1469

Qy       1495 AGGAAGGTTTTGGGCAGGACGTAGCTGGAAGCCTGAGCTTGTCACCCATGGGGATGGGGA 1554
Db       1470 AGGACAATTTTGCAAAAGAAGTAGCTGGAAGCC--------------ATGGGACTGGCG 1514
```

```
Qy       1555 GAGCCCTGTTTGAGGGCGGCTGATGGTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAA 1614
Db       1515 GGAGCCTGTTTGGGGGATCTGGATGGTTGACTCC--------TAGGAGTCAAGTTCAGCA 1566

Qy       1615 TCTTCCCAGTGGCCTGTAGAGTTGCCTCCTGACCACTAGAGGGCGCGCCCACACAGCATT 1674
Db       1567 TCTTCGCCGTGG-CTGCAGAGCTGCCTGATGGGCACTAGAGGGCACGCCAGCCCCACACT 1625

Qy       1675 ACCTGGGTCTGCCTTTCCTAGGACAACCCCACCCAGTACAGCCCTGTGCCTGGTGTGTCC 1734
Db       1626 CCCTGGGTCTGGCTTCCTCCCGCAACCTCACTCTAGTAGAGCCTGTG------------ 1672

Qy       1735 ACCCTGCTTACTAGTTCTTTGGGTTTCATGGAATTTACAAGCTTCTAAAGGAGCAGAGTG 1794
Db       1673 --CCTGCCTACTAGCGCTCTGGGGTTCGGAGAGTTTGGGAATTTCTCA--GAGCCAACTG 1728

Qy       1795 GCTCAGATTGGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGT--------GT 1846
Db       1729 GCTCAGGCTTGGGAAGGCTGGCTGCTGCCCTCAGCTCCGCCTCATCAGCTATGTGAAGGG 1788

Qy       1847 GTGTGGAGTATTTGTGAATCAAAGGAGAGGTTTGGCCTAGTGCCCAGTCTTTTAACTTAG 1906
Db       1789 GTGTGTATGGAGTGATCCTGCCGCCCCCTCCCTGGGCTGGTCCAGAGATCTCAAACTCCG 1848

Qy       1907 ATGCCCTCAGGGCCGGGTGGGTTATAAAA----ATAAAGTAGGCCTTTGAGCTGTGAGGC 1962
Db       1849 ATGCCCCTGGGGCCACGTATGTTGTGTAAATGGATGAAACAGGCCCTTGAGTTGGGAGCC 1908

Qy       1963 CTTTGGGACTTTAATTTTTCCCACTATTCCTGGAGATGGGACATAGAGAGACATTGCTTT 2022
Db       1909 TGCTTCACTTTGA--CTTTCCCACTGTTGCTGGAGA-------------CAAAGACATC 1952

Qy       2023 GTGCTGAGAAATACTTGCATGATTGAGTCTGAGTCGCTAAGGGCAACTGGCCTTGAGTGA 2082
Db       1953 GTGATGAGAGAAAGTTCGCACAATCT-----AGTCGGTAACAGCCACTTTCCTTGAGACC 2007

Qy       2083 CATCAAGGGGTGGTGGGGACTGTGGCAAACCACAGATTCCCACCTGAAATTGGTGGCTGT 2142
Db       2008 AAGAGAGTGCGGTGGGGATGGGGGGGAGAGCACGGGTCCCCGTCTGACAGTGGCCGCTGC 2067

Qy       2143 C--CTTCCGTTGGGGCTAATGGCTGTACAGCGAGAATGTAGGTAGGTCTGTCTAATGGGA 2200
Db       2068 CATATTCAGGTGTAGCTAATTGCTCTGGTGTGGGAATGCAG--------GCCTAATGACA 2119

Qy       2201 GAAGTCTGGAGAAGCCAAGAAGCTAGATTTTTCATGTGAACTATCCCGAGTTTTAAGTTG 2260
Db       2120 GAAATCTGGAGAAGCCAGAAATACAGATTTGT-ATGTGAGATGTCCTGATTTTTTAAGTT 2178

Qy       2261 TTTGCAGCTAATGAGAAAAACCTCTTAAACCCTGATAGTCAAAAGGTGTGGGGGCCATCT 2320
Db       2179 GTTGGCAGAAATTAATTCAGAAATCAAATCTGCAGGCCAAACAAGGTGCAGGACCCAGCT 2238

Qy       2321 TTGACACCTCCCCCACCATAGGTCCCTCAGGGACAGTGCCCCATGGGAGCCCTGGTGAGT 2380
Db       2239 TTGGCCCCATGCCC-CTGTAGGTCCCTCTGGGACAGTCACCGCTGGG-GTCCTGGCTGCT 2296

Qy       2381 CCACGGTTCAGGAATGCTGGGAACTGCTGCAGGTGGGC 2418
Db       2297 CTGTCATTGAGGGATGCTGGGCACTGCTGCCGGGTGGC 2334
```

```
RESULT 3
US-09-934-268-1
; Sequence 1, Application US/09934268
; Patent No. US20020172986A1
; GENERAL INFORMATION:
;  APPLICANT: Leiby, Kevin R.
;  TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY
;  TITLE OF INVENTION:  MEMBER AND USES THEREOF
;  FILE REFERENCE: 10448-079001
;  CURRENT APPLICATION NUMBER: US/09/934,268
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3184
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (168)...(977)
US-09-934-268-1

  Query Match            31.1%;  Score 790.4;  DB 10;  Length 3184;
  Best Local Similarity  92.3%;  Pred. No. 7.2e-220;
  Matches  855;  Conservative    0;  Mismatches   66;  Indels    5;  Gaps    2;

Qy          4 CGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTG 63
Db         52 CGCATGAAGCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCG 111

Qy         64 GAGGGCCGATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGC 122
Db        112 GGGCGCTGGGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGC 171

Qy        123 CGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 182
Db        172 CCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 231

Qy        183 TTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGG 242
Db        232 TTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGG 291

Qy        243 CCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCG 302
Db        292 CCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTG 351

Qy        303 ACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCT 362
Db        352 ACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCT 411

Qy        363 GCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCC 422
Db        412 GCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTC 471

Qy        423 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTC 482
Db        472 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTC 531

Qy        483 GAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACC 542
Db        532 GAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACC 591

Qy        543 TCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTA 602
Db        592 TCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGCCCCA 651

Qy        603 ATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCT 662
Db        652 ATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCT 711

Qy        663 GCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGT 718
Db        712 GCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGT 771

Qy        719 GTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778
Db        772 GTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAG 831

Qy        779 GCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCT 838
Db        832 GCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCT 891

Qy        839 TTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTG 898
```

```
Db        892 TCATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTG 951

Qy        899 ACATCAAGGCAGTGAAGGCCAACTGG 924
              |||||||||||||||| |||||||||
Db        952 ACATCAAGGCAGTGAAAGCCAACTGG 977
```

RESULT 4
US-10-162-435-35
; Sequence 35, Application US/10162435
; Publication No. US20030096305A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Curtis , Rory A. J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Leiby , Kevin R.
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND
; TITLE OF INVENTION: CELL SURFACE PROTEIN FAMILY MEMBERS
; FILE REFERENCE: 10448-189001
; CURRENT APPLICATION NUMBER: US/10/162,435
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/836,499
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: PCT/US01/12420
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 60/197,507
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/891,008
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: PCT/US01/19963
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/214,220
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/860,868
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/16013
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,674
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/886,429
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: PCT/US01/20055
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/213,963
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 10/041,406
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: PCT/US02/00275
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/260,286
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/934,268
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/41811
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,612
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 3184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (168)...(977)
US-10-162-435-35

Query Match 31.1%; Score 790.4; DB 14; Length 3184;
Best Local Similarity 92.3%; Pred. No. 7.2e-220;

Matches 855; Conservative 0; Mismatches 66; Indels 5; Gaps 2;

```
Qy          4 CGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTG 63
              ||| |  ||||||||||| ||||||||||||||||||||||||||||||||||||||| |
Db         52 CGCATGAAGCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCG 111

Qy         64 GAGGGCCGATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGC 122
              | | || |  || ||| |||||||||||||| |||||||||||||| ||| ||||||||||
Db        112 GGGCGCTGGGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGC 171

Qy        123 CGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 182
              | |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||
Db        172 CCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 231

Qy        183 TTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGG 242
              ||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db        232 TTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGG 291

Qy        243 CCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCG 302
              |||||||||||||||| |||||||| |||||||| |||||||| |||||||||||||| |
Db        292 CCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTG 351

Qy        303 ACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCT 362
              |||| |||||||||||||| || |||||||||||||||| |||||||||||||| ||||
Db        352 ACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCT 411

Qy        363 GCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCC 422
              |||| ||||| |||||||||||||| |||||||||||||||||||| |||||||| || |
Db        412 GCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTC 471

Qy        423 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTC 482
              |||||||||||||||||||||||||||||||| ||||||| || ||||||||||||||||
Db        472 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTC 531

Qy        483 GAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACC 542
              |||||||||||| ||||||| |||||||||||||||||||| ||||| ||||||||||
Db        532 GAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACC 591

Qy        543 TCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTA 602
              |||||||||||||||||||||||||||||||||||| |||||||||| |||||||||| || |
Db        592 TCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGCCCCA 651

Qy        603 ATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCT 662
              |||||||||||||||||||||| |||||||||||||||||||||| ||||| |||||||
Db        652 ATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCT 711

Qy        663 GCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGT 718
              |||||||||||||||||||| |||||||||||||    ||||||||||||||||||||
Db        712 GCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGT 771

Qy        719 GTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778
              ||||||| || |||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db        772 GTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAG 831

Qy        779 GCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCT 838
              |||| |||||||||||||||||||||||||||||||||||||| |||||||| || ||||
Db        832 GCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCT 891

Qy        839 TTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTG 898
              |  |||||||||| |||||||||||||||| |||||||||||||||||||| |||||||
Db        892 TCATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTG 951

Qy        899 ACATCAAGGCAGTGAAGGCCAACTGG 924
              |||||||||||||||| |||||||||
Db        952 ACATCAAGGCAGTGAAAGCCAACTGG 977
```

RESULT 5
US-10-103-196-4
; Sequence 4, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.

```
;  TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE: PT008P1
;  CURRENT APPLICATION NUMBER: US/10/103,196
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/707,936
;  PRIOR FILING DATE: 2000-11-08
;  PRIOR APPLICATION NUMBER: PCT/US00/13504
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: 60/178,770
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/149,447
;  PRIOR FILING DATE: 1999-08-18
;  PRIOR APPLICATION NUMBER: 60/138,573
;  PRIOR FILING DATE: 1999-06-11
;  PRIOR APPLICATION NUMBER: 60/137,797
;  PRIOR FILING DATE: 1999-06-03
;  PRIOR APPLICATION NUMBER: 60/135,122
;  PRIOR FILING DATE: 1999-05-19
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1126
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-103-196-4

  Query Match             31.0%;  Score 787.2;  DB 14;  Length 1126;
  Best Local Similarity   92.6%;  Pred. No. 3.6e-219;
  Matches  850;  Conservative    0;  Mismatches   63;  Indels    5;  Gaps    2;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||||||||||||| || | || |
Db         49 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 108

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||||| ||| |||||||||||| ||||||
Db        109 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 168

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        169 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 228

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              |||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        229 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 288

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
              |||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| |||
Db        289 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTG 348

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
              |||||||||| || |||||||||||||||| |||||||||| |||||||| |||
Db        349 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 408

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
              || |||||||||||||||| ||||||| || |||||||| || |||||||||
Db        409 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 468

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
              |||||||||||||||||||||| ||||||| || |||||||||||||||||||||||||
Db        469 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 528

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
              |||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||||
Db        529 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 588

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
              |||||||||||||||||||||| |||||||||| |||||||||| || |||||||||
Db        589 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGCCCCAATGACTGG 648

Qy        611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
              |||||||||||||| |||||||||||||||||||||| ||||| ||||||||||||||||
Db        649 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 708

Qy        671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
              |||||||||||||| |||||||||||||    |||||||||||||||||||||||||||
Db        709 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTAC 768

Qy        727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTG 786
              || |||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||| |||
Db        769 GACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTG 828

Qy        787 GGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGC 846
              |||||||||||||| |||||||||||||||||||| |||||||| || |||||  |||||
Db        829 GGCCAGTTTGAGAAATGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGC 888

Qy        847 ATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAG 906
              |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db        889 ATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAAG 948

Qy        907 GCAGTGAAGGCCAACTGG 924
              |||||||| |||||||||
Db        949 GCAGTGAAAGCCAACTGG 966


RESULT 6
US-09-934-268-3
; Sequence 3, Application US/09934268
; Patent No. US20020172986A1
; GENERAL INFORMATION:
;  APPLICANT: Leiby, Kevin R.
;  TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY
;  TITLE OF INVENTION:  MEMBER AND USES THEREOF
;  FILE REFERENCE: 10448-079001
;  CURRENT APPLICATION NUMBER: US/09/934,268
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 813
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-934-268-3

  Query Match             28.0%;  Score 710.4;  DB 10;  Length 813;
  Best Local Similarity   93.2%;  Pred. No. 8.7e-197;
  Matches  755;  Conservative    0;  Mismatches   51;  Indels    4;  Gaps    1;

Qy        119 ATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTC 178
              ||||| |||||||||||||| |||||||||| |||||||||||||||||||||||||||
Db          1 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 60

Qy        179 CTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTC 238
              |||||||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db         61 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 120

Qy        239 TGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGC 298
              |||||||||||||||||||| |||||||| |||||||| |||||||| ||||||||||||
Db        121 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 180

Qy        299 CTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCC 358
              || ||||| |||||||||||||| || |||||||||||||||| ||||||||||||||||
Db        181 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 240

Qy        359 GGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTT 418
              |||||||| ||||| |||||||||||||| |||||||||||||||||| ||||||||
Db        241 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 300

Qy        419 GGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGG 478
              || |||||||||||||||||||||||||||||||||| ||||||| || ||||||||||||
Db        301 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 360

Qy        479 ATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATT 538
              |||||||||||||||||  ||||||| |||||||||||||||||||| ||||| ||||||
```

```
Db        361 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 420

Qy        539 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGG 598
              ||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db        421 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 480

Qy        599 CCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAG 658
              || |||||||||||||||||||||| |||||||||||||||||||||| ||||| |||
Db        481 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 540

Qy        659 CGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACC 714
              |||||||||||||||||||||||||| |||||||||||||    ||||||||||||||
Db        541 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 600

Qy        715 CAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACC 774
              |||||||||| || ||||||||||||||||||||||||||||||||| ||| ||||||
Db        601 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 660

Qy        775 AAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGG 834
              |||||||| |||||||||||||||||||||||||||||||||||||| |||||||| ||
Db        661 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGA 720

Qy        835 GTCTTTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTG 894
              |||||  |||||||||| |||||||||||||||||| |||||||||||||||||||| |||
Db        721 GTCTTCATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTG 780

Qy        895 AGTGACATCAAGGCAGTGAAGGCCAACTGG 924
              |||||||||||||||||||| |||||||||
Db        781 AGTGACATCAAGGCAGTGAAAGCCAACTGG 810


RESULT 7
US-10-162-435-37
; Sequence 37, Application US/10162435
; Publication No. US20030096305A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Curtis , Rory A. J.
;  APPLICANT:  Kapeller-Libermann, Rosana
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Leiby , Kevin R.
;  TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND
;  TITLE OF INVENTION:  CELL SURFACE PROTEIN FAMILY MEMBERS
;  FILE REFERENCE: 10448-189001
;  CURRENT APPLICATION NUMBER: US/10/162,435
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US 09/836,499
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: PCT/US01/12420
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: US 60/197,507
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: US 09/891,008
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: PCT/US01/19963
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/214,220
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 09/860,868
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: PCT/US01/16013
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,674
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/886,429
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: PCT/US01/20055
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/213,963
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 10/041,406
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: PCT/US02/00275
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/260,286
;  PRIOR FILING DATE: 2001-01-08
;  PRIOR APPLICATION NUMBER: US 09/934,268
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/41811
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
;   LENGTH: 813
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-162-435-37

  Query Match             28.0%;  Score 710.4;  DB 14;  Length 813;
  Best Local Similarity   93.2%;  Pred. No. 8.7e-197;
  Matches  755;  Conservative    0;  Mismatches   51;  Indels    4;  Gaps    1;

Qy        119 ATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTC 178
              ||||| |||||||||||||| |||||||||| |||||||||||||||||||||||||||
Db          1 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 60

Qy        179 CTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTC 238
              ||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db         61 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 120

Qy        239 TGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGC 298
              |||||||||||||||||||| |||||||| |||||||| |||||||| ||||||||||||
Db        121 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 180

Qy        299 CTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCC 358
              || ||||| |||||||||||||| || |||||||||||||||| ||||||||||||||
Db        181 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 240

Qy        359 GGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTT 418
              |||||||| ||||| |||||||||||||| |||||||||||||||||||| ||||||||
Db        241 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 300

Qy        419 GGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGG 478
              || |||||||||||||||||||||||||||||||||| ||||||| || ||||||||||||
Db        301 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 360

Qy        479 ATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATT 538
              |||||||||||||||  ||||||| |||||||||||||||||||| ||||| ||||||
Db        361 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 420

Qy        539 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGG 598
              |||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
Db        421 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 480

Qy        599 CCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAG 658
              || |||||||||||||||||||||| |||||||||||||||||||||| ||||| |||
Db        481 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 540

Qy        659 CGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACC 714
              |||||||||||||||||||||||||| |||||||||||||    ||||||||||||||
Db        541 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 600

Qy        715 CAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACC 774
              |||||||||| || ||||||||||||||||||||||||||||||||| ||| ||||||
Db        601 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 660

Qy        775 AAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGG 834
              |||||||| |||||||||||||||||||||||||||||||||||||| |||||||| ||
Db        661 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGA 720

Qy        835 GTCTTTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTG 894
              |||||  |||||||||| |||||||||||||||||| |||||||||||||||||||| |||
Db        721 GTCTTCATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTG 780
```

```
Qy        895 AGTGACATCAAGGCAGTGAAGGCCAACTGG 924
              |||||||||||||||||||| |||||||||
Db        781 AGTGACATCAAGGCAGTGAAAGCCAACTGG 810
```

RESULT 8
US-09-925-299-205
; Sequence 205, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 205
; LENGTH: 1655
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1548)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1559)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1564)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1623)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1643)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-205

Query Match 17.4%; Score 441.6; DB 9; Length 1655;
Best Local Similarity 73.2%; Pred. No. 4.7e-118;
Matches 581; Conservative 0; Mismatches 209; Indels 4; Gaps 1;

```
Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        287 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 346

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||      || || |  || || || ||||| |||  ||| |
Db        347 TCATATTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 406

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | || ||||| |||||||||||| |  | |||||||| || ||| | ||||| || ||||
Db        407 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 466

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | |  |||| ||||| || ||||| || || |
Db        467 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGCGC 526

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCC 434
              | |||||| |||||||||| |||||||||| || |||||||| ||  | || |||||||
Db        527 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATTATTTTCTTCC 586

Qy        435 TGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCA 494
              |||||||  |  | ||  |  | || || || ||||| ||||||||   |||||||||
Db        587 TGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTGT 646

Qy        495 ATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCA 554
              ||||||| || ||||||||| |||  || |||||||||||||||||  | || |||||||
Db        647 ATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTCA 706

Qy        555 TTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACC 614
              | |||||  | ||||||||||||    ||||| || ||     ||  || |||| |||||||
Db        707 TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAACC 766

Qy        615 TCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCT 674
              | ||||| || ||||| ||||| || |  ||  | || |||||||| || || || || |
Db        767 TAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCAT 826

Qy        675 TCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATG 730
              ||||||||||     |  || || ||    |||||| ||||||| ||||||||||||||||
Db        827 TCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGATG 886

Qy        731 TCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCC 790
              | |||  ||||  ||  | || ||||||      ||  |||| ||||||||||||  ||
Db        887 CCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCCC 946

Qy        791 AGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCG 850
              |||||||||||||| ||||||||||  | | | | || |||||  | ||  | ||||| |
Db        947 AGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATTG 1006

Qy        851 CTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAG 910
              |  | || ||||| ||||| || ||||||||||||||  | || || || ||| | || |
Db       1007 CATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCTG 1066

```
Qy        911 TGAAGGCCAACTGG 924
              | | ||| | ||||
Db       1067 TCAGGGCGAGCTGG 1080
```

RESULT 9
US-09-925-299-205
; Sequence 205, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 205
; LENGTH: 1655
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1548)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1559)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1564)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1623)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1643)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-205

Query Match 17.4%; Score 441.6; DB 11; Length 1655;
Best Local Similarity 73.2%; Pred. No. 4.7e-118;
Matches 581; Conservative 0; Mismatches 209; Indels 4; Gaps 1;
```

```
Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
Db        287 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 346

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
Db        347 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 406

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
Db        407 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 466

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
Db        467 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGCGC 526

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCC 434
Db        527 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATTATTTTCTTCC 586

Qy        435 TGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCA 494
Db        587 TGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTGT 646

Qy        495 ATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCA 554
Db        647 ATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTCA 706

Qy        555 TTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACC 614
Db        707 TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAACC 766

Qy        615 TCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCT 674
Db        767 TAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCAT 826

Qy        675 TCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATG 730
Db        827 TCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGATG 886

Qy        731 TCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCC 790
Db        887 CCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCCC 946

Qy        791 AGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCG 850
Db        947 AGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATTG 1006

Qy        851 CTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAG 910
Db       1007 CATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCTG 1066

Qy        911 TGAAGGCCAACTGG 924
Db       1067 TCAGGGCGAGCTGG 1080
```

RESULT 10
US-10-103-196-5
; Sequence 5, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT008P1
; CURRENT APPLICATION NUMBER: US/10/103,196
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/707,936
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/13504
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/178,770
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/149,447
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/138,573
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/137,797
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/135,122
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-196-5

Query Match 16.5%; Score 418.6; DB 14; Length 1174;
Best Local Similarity 73.0%; Pred. No. 2.1e-111;
Matches 580; Conservative 0; Mismatches 209; Indels 6; Gaps 3;

```
Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
Db        103 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 162

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
Db        163 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 222

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
Db        223 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 282

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
Db        283 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGC-G 341

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC-TTC 433
Db        342 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATTATTTTCTTTC 401

Qy        434 CTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTC 493
Db        402 CTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTG 461

Qy        494 AATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTC 553
Db        462 TATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTC 521

Qy        554 ATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAAC 613
Db        522 ATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAAC 581

Qy        614 CTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCC 673
Db        582 CTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCA 641

Qy        674 TTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGAT 729
Db        642 TTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGAT 701

Qy        730 GTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGC 789
Db        702 GCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCC 761
```

```
Qy        790 CAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATC 849
              ||||||||||||||||| |||||||||| | | | | || ||||| | ||  | |||||
Db        762 CAGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATT 821

Qy        850 GCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCA 909
              ||  | || ||||| ||||| || ||||||||||||||| | || || || ||| | ||
Db        822 GCATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCT 881

Qy        910 GTGAAGGCCAACTGG 924
              || | ||| | ||||
Db        882 GTCAGGGCGAGCTGG 896


RESULT 11
US-10-103-196-10
; Sequence 10, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE: PT008P1
;  CURRENT APPLICATION NUMBER: US/10/103,196
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/707,936
;  PRIOR FILING DATE: 2000-11-08
;  PRIOR APPLICATION NUMBER: PCT/US00/13504
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: 60/178,770
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/149,447
;  PRIOR FILING DATE: 1999-08-18
;  PRIOR APPLICATION NUMBER: 60/138,573
;  PRIOR FILING DATE: 1999-06-11
;  PRIOR APPLICATION NUMBER: 60/137,797
;  PRIOR FILING DATE: 1999-06-03
;  PRIOR APPLICATION NUMBER: 60/135,122
;  PRIOR FILING DATE: 1999-05-19
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 1178
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-103-196-10

  Query Match             15.8%;  Score 402.2;  DB 14;  Length 1178;
  Best Local Similarity   71.6%;  Pred. No. 1.3e-106;
  Matches  571;  Conservative    0;  Mismatches  218;  Indels    8;  Gaps    3;

Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        101 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 160

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||      || || |  || || || ||||| |||  ||| |
Db        161 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 220

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | || ||||| |||||||||||| |  | |||||||| || ||| | ||||| || ||||
Db        221 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 280

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | |  |||| ||||| || ||||| || ||
Db        281 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGC-G 339

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTC---AGTGTTCCTTGGCCTCATCTTCT 431
              | ||||| |||||||||||| |||||||||||| ||    ||    |  |   |  | |  |
Db        340 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTCGTGTTCCTCGGGAATTATTTTCTT 399

Qy        432 TCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGC 491
              ||||||||||  |  | ||  |  | || || || ||||| ||||||||   ||||||||
Db        400 TCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGC 459
```

```
Qy        492 TCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACC 551
              |  |||||| || ||||||||| |||  || ||||||||||||||||||  | || ||||
Db        460 TGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACC 519

Qy        552 TCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGA 611
              |||| |||||  | |||||||||||  |||| || ||    || || |||| ||||
Db        520 TCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGA 579

Qy        612 ACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGC 671
              |||| ||||| || ||||| ||||| || |  ||  | || |||||||| || || | |
Db        580 ACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTC 639

Qy        672 CCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATG 727
              | ||||||||||     |  || || ||    |||||| ||||||| ||||||||||||
Db        640 CATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATG 699

Qy        728 ATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGG 787
              ||| | |||  ||||  ||  | || ||||||       ||  |||| ||||||||||||
Db        700 ATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGC 759

Qy        788 GCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCA 847
              |||||||||||||||| ||||||||||  | | | | || |||||  | ||  | ||||
Db        760 CCCAGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCA 819

Qy        848 TCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGG 907
              | ||  | || ||||| ||||| || ||||||||||||||  | || || || ||| | |
Db        820 TTGCATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAG 879

Qy        908 CAGTGAAGGCCAACTGG 924
              | || | ||| | ||||
Db        880 CTGTCAGGGCGAGCTGG 896


RESULT 12
US-09-822-846-126
; Sequence 126, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Steininger II, Robert J.
;  APPLICANT:  Bowman, Michael R.
;  APPLICANT:  Spaulding, Vikki
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6400
;  CURRENT APPLICATION NUMBER: US/09/822,846
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,605
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 629
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 126
;   LENGTH: 1988
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-846-126

  Query Match             15.4%;  Score 390.8;  DB 11;  Length 1988;
  Best Local Similarity   74.3%;  Pred. No. 3.7e-103;
```

```
  Matches  508;  Conservative    0;  Mismatches  172;  Indels    4;  Gaps    1;

Qy        245 TGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGAC 304
              ||| ||| || || ||||| |||||||||||| |  | |||||||| || ||| | |||
Db          2 TGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGAC 61

Qy        305 CCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGC 364
              || || ||||| ||  | ||||| ||||| || |||| | |  |||| ||||| || |||
Db         62 CCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGC 121

Qy        365 ATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTC 424
              || || || || ||||| |||||||||||| |||||||||| || |||||||| ||   |
Db        122 ATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATT 181

Qy        425 ATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGA 484
              || ||||||||||||||  |  | ||  |  | || || || ||||| ||||||||   |
Db        182 ATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAA 241

Qy        485 GACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTC 544
              |||||||  |||||| || |||||||| |||  || ||||||||||||||||||   |
Db        242 GACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTG 301

Qy        545 CAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAAT 604
              || |||||||| |||||  | ||||||||||||    ||||| || ||     || || ||
Db        302 CAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGAT 361

Qy        605 GACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGC 664
              || |||||||| ||||| || ||||| ||||| || |  ||  | || |||||||| ||
Db        362 GATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGT 421

Qy        665 GGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGT 720
              || || || ||||||||||     | || || ||    |||||| ||||||| ||||||
Db        422 GGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGT 481

Qy        721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGC 780
              |||||||||| | |||  |||| ||  | || ||||||     ||  |||| ||||||
Db        482 GGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGC 541

Qy        781 TGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTT 840
              ||||||  |||||||||||||||| |||||||||  | | | | || |||||  | ||
Db        542 TGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTC 601

Qy        841 GTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGAC 900
              | |||||  ||  | || ||||| ||||| || |||||||||||||  | || || ||
Db        602 ATAGGCATTGCATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGAT 661

Qy        901 ATCAAGGCAGTGAAGGCCAACTGG 924
              ||| | || || | ||| | ||||
Db        662 ATCGAAGCTGTCAGGGCGAGCTGG 685


RESULT 13
US-09-892-877-20
; Sequence 20, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et. al.
;  TITLE OF INVENTION: 97 Human secreted proteins
;  FILE REFERENCE: PZ028P1
;  CURRENT APPLICATION NUMBER: US/09/892,877
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
;  NUMBER OF SEQ ID NOS: 461
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 2672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (28)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (47)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (52)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (93)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-892-877-20

  Query Match             14.1%;  Score 357.4;  DB 11;  Length 2672;
  Best Local Similarity   64.3%;  Pred. No. 2.5e-93;
  Matches  564;  Conservative    3;  Mismatches  304;  Indels    6;  Gaps    2;

Qy         48 CCCAGGGCGGCCCCGTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGC 107
              ||| |||| ||| |        |    | |||| || | ||| | ||||    | |||:||
Db         99 CCCGGGGCTGCCAGGAATTTCCGAGCCGGGGCCGCGCCGCCGCTGCCCGCCGCCGCGSGC 158

Qy        108 GGCTGCT--CACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTG 165
              || | :|   |  | |  |  | | |  | |   |  || ||||| |||||||
Db        159 GGATTYTGCTTCTCAGAAGATGCACTATTATAGATACTCTAACGCCAAGGTCAGCTGCTG 218

Qy        166 CGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCT 225
                || ||| |||| ||  ||| |||||  | |||||: ||  ||||  | |||||
Db        219 GTACAAGTACCTCCTTTTCAGCTACAACATCATCTTCTGRTTGGCTGGAGTTGTCTTCCT 278

Qy        226 GGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGAC 285
              |  |||| || ||||| ||| | || |||||||| || ||| || ||  |    ||||
Db        279 TGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCGACCTCACCAAAGTGAC 338

Qy        286 CGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGT 345
              |  ||  ||  |||||||||||  |||| |  |  |||| || |  || |||| |
Db        339 CCGGATGCATGGAATCGACCCTGTGGTGCTGGTCCTGATGGTGGGCGTGGTGATGTTCAC 398

Qy        346 GCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTT 405
              ||||| || |||||||| | |||||||| |||||||| |  | | ||||||| |||||
Db        399 CCTGGGGTTCGCCGGCTGCGTGGGGGCTCTGCGGGAGAATATCTGCTTGCTCAACTTTTT 458

Qy        406 CTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGT 465
              ||  |   | | |  |||||||||||||||||||||||     |   | ||||||| |
Db        459 CTGTGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGGCCGTGCTGGCCTTCCT 518

Qy        466 ATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTA 525
              ||| |||||||| |  | |||| | ||      ||| |  | | |||| ||||| ||||
Db        519 GTTCCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGAGCAACATCAAGTCCTA 578

Qy        526 TCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTG 585
              ||||| || || || || || |||||||| ||||     |||| ||  |       |||||
Db        579 CCGGGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGAAAGCTAACCAGTGCTG 638

Qy        586 CGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAA 645
              || ||     || ||| | |||||| |||||||  |||| ||||| ||||  |     |
Db        639 TGGCGCATATGGCCCTGAAGACTGGGACCTCAACGTCTACTTCAATTGCAGCGGTGCCAG 698

Qy        646 CCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCG----A 701
              |   |||||||||   |||||||| ||||||||||||||| ||    || ||||||    |
Db        699 CTACAGCCGAGAGAAGTGCGGGGTCCCCTTCTCCTGCTGCGTGCCAGATCCTGCGCAAAA 758

Qy        702 TGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGG 761
              ||  | ||||| |||||||| |||||||| || |  | ||| ||    || | | |
Db        759 AGTTGTGAACACACAGTGTGGATATGATGTCAGGATTCAGCTGAAGAGCAAGTGGGATGA 818

Qy        762 CTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGAT 821
              |||||   ||| ||||||||  |       | | || |  ||||| | |   ||| |
Db        819 GTCCATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGCTCCCGCGGAACATTTA 878

Qy        822 CGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGC 881
              | | |||||||| |||||  | | |||| | ||  | ||||| ||||| |||| ||||||
Db        879 CATTGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATATTTGGCATCTTCCTGGC 938
```

```
Qy        882 CCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCC 918
              ||   ||   |    |||||| ||||||||||||||
Db        939 AAGGACGCTGATCTCAGACATCGAGGCAGTGAAGGCC 975
```

```
RESULT 14
US-09-948-783-20
; Sequence 20, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et. al.
;  TITLE OF INVENTION: 97 Human secreted proteins
;  FILE REFERENCE: PZ028P2
;  CURRENT APPLICATION NUMBER: US/09/948,783
;  CURRENT FILING DATE:  2001-09-10
;  PRIOR APPLICATION NUMBER: 60/231,846
;  PRIOR FILING DATE: 2000-09-11
;  PRIOR APPLICATION NUMBER: 09/892,877
;  PRIOR FILING DATE: 2001-06-28
;  PRIOR APPLICATION NUMBER: 09/437,658
;  PRIOR FILING DATE: 1999-11-10
;  PRIOR APPLICATION NUMBER: PCT/US99/09847
;  PRIOR FILING DATE: 1999-05-06
;  PRIOR APPLICATION NUMBER: 60/085,093
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,094
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,105
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,180
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,927
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,906
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,924
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,922
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,921
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,923
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,925
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,928
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,920
;  PRIOR FILING DATE: 1998-05-18
;  NUMBER OF SEQ ID NOS: 465
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 2672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (28)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (47)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (52)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (93)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-948-783-20
```

```
  Query Match             14.1%;  Score 357.4;  DB 11;  Length 2672;
  Best Local Similarity   64.3%;  Pred. No. 2.5e-93;
  Matches  564;  Conservative    3;  Mismatches  304;  Indels    6;  Gaps    2;

Qy         48 CCCAGGGCGGCCCGTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGC 107
              ||| |||| ||| |         |     | |||| || | ||| | |||| | |||:||
Db         99 CCCGGGGCTGCCAGGAATTTCCGAGCCGGGGCCGCGCCGCCGCTGCCCGCCGCCGCGSGC 158

Qy        108 GGCTGCT--CACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTG 165
              || | :|    |   | |    | | |    | |    | || ||||| |||||||
Db        159 GGATTYTGCTTCTCAGAAGATGCACTATTATAGATACTCTAACGCCAAGGTCAGCTGCTG 218

Qy        166 CGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCT 225
                || ||| |||| ||  ||| ||||||  | |||||: ||   ||||   | |||||
Db        219 GTACAAGTACCTCCTTTTCAGCTACAACATCATCTTCTGRTTGGCTGGAGTTGTCTTCCT 278

Qy        226 GGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGAC 285
              |   |||| || ||||| ||| | || |||||||| || ||| || || | .  ||||
Db        279 TGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCGACCTCACCAAAGTGAC 338

Qy        286 CGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGT 345
              |    ||   ||  ||||||||||  |||| |  |  |||| || |  || |||| |
Db        339 CCGGATGCATGGAATCGACCCTGTGGTGCTGGTCCTGATGGTGGGCGTGGTGATGTTCAC 398

Qy        346 GCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTT 405
              ||||| || |||||||||  | |||||||| ||||||||  |  | | |||||||  |||||
Db        399 CCTGGGGTTCGCCGGCTGCGTGGGGGCTCTGCGGGAGAATATCTGCTTGCTCAACTTTTT 458

Qy        406 CTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGT 465
              ||  |    | | |  ||||||||||||||||||||||||     |    |  |||||||| |
Db        459 CTGTGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGGCCGTGCTGGCCTTCCT 518

Qy        466 ATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTA 525
              |||  |||||||| |  | |||| | ||       ||| |  | | |||| ||||| ||||
Db        519 GTTCCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGAGCAACATCAAGTCCTA 578

Qy        526 TCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTG 585
              |||||  || || || || || ||||||||  ||||     |||| ||   |      |||||
Db        579 CCGGGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGAAAGCTAACCAGTGCTG 638

Qy        586 CGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAA 645
              || ||    || ||| | |||||| |||||||  |||| |||||  ||||   |      |
Db        639 TGGCGCATATGGCCCTGAAGACTGGGACCTCAACGTCTACTTCAATTGCAGCGGTGCCAG 698

Qy        646 CCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCG----A 701
              |   |||||||||    |||||||| |||||||||||||| ||     || ||||||    |
Db        699 CTACAGCCGAGAGAAGTGCGGGGTCCCCTTCTCCTGCTGCGTGCCAGATCCTGCGCAAAA 758

Qy        702 TGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGG 761
              ||  | ||||| |||||||| |||||||| || |  | ||| ||    ||  | | |
Db        759 AGTTGTGAACACACAGTGTGGATATGATGTCAGGATTCAGCTGAAGAGCAAGTGGGATGA 818

Qy        762 CTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGAT 821
              |||||   ||| |||||||| |       | | || |  ||||| | |   ||| |
Db        819 GTCCATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGCTCCCGCGGAACATTTA 878

Qy        822 CGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGC 881
              | | |||||||| |||||  | | |||| | ||  | ||||| ||||| |||| ||||||
Db        879 CATTGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATATTTGGCATCTTCCTGGC 938

Qy        882 CCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCC 918
                ||   ||   |    |||||| ||||||||||||||
Db        939 AAGGACGCTGATCTCAGACATCGAGGCAGTGAAGGCC 975
```

```
RESULT 15
US-09-796-753-125
; Sequence 125, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
;  APPLICANT:  McCarthy, Sean A.
;  TITLE OF INVENTION:  SECRETED PROTEINS AND USES THEREOF
```

```
;  FILE REFERENCE:  7853-227-999
;  CURRENT APPLICATION NUMBER: US/09/796,753
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER:  09/183,175
;  PRIOR FILING DATE:  1998-10-30
;  PRIOR APPLICATION NUMBER:  09/223,094
;  PRIOR FILING DATE:  1998-12-30
;  PRIOR APPLICATION NUMBER:  09/223,546
;  PRIOR FILING DATE:  1998-12-30
;  PRIOR APPLICATION NUMBER:  09/224,246
;  PRIOR FILING DATE:  1998-12-30
;  PRIOR APPLICATION NUMBER:  09/259,388
;  PRIOR FILING DATE:  1999-02-26
;  PRIOR APPLICATION NUMBER:  60/122,458
;  PRIOR FILING DATE:  1999-03-01
;  PRIOR APPLICATION NUMBER:  09/312,359
;  PRIOR FILING DATE:  1999-05-14
;  PRIOR APPLICATION NUMBER:  09/336,536
;  PRIOR FILING DATE:  1999-06-18
;  PRIOR APPLICATION NUMBER:  09/342,687
;  PRIOR FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER:  09/345,464
;  PRIOR FILING DATE:  1999-06-30
;  PRIOR APPLICATION NUMBER:  09/365,164
;  PRIOR FILING DATE:  1999-07-30
;  PRIOR APPLICATION NUMBER:  09/399,723
;  PRIOR FILING DATE:  1999-09-20
;  PRIOR APPLICATION NUMBER:  09/409,634
;  PRIOR FILING DATE:  1999-09-30
;  PRIOR APPLICATION NUMBER:  09/471,179
;  PRIOR FILING DATE:  1999-12-23
;  PRIOR APPLICATION NUMBER:  09/474,071
;  PRIOR FILING DATE:  1999-12-29
;  PRIOR APPLICATION NUMBER:  09/474,072
;  PRIOR FILING DATE:  1999-12-29
;  PRIOR APPLICATION NUMBER:  09/514,010
;  PRIOR FILING DATE:  2000-02-25
;  PRIOR APPLICATION NUMBER:  09/516,745
;  PRIOR FILING DATE:  2000-03-01
;  PRIOR APPLICATION NUMBER:  09/572,002
;  PRIOR FILING DATE:  2000-05-14
;  PRIOR APPLICATION NUMBER:  09/597,993
;  PRIOR FILING DATE:  2000-06-19
;  PRIOR APPLICATION NUMBER:  09/599,596
;  PRIOR FILING DATE:  2000-06-22
;  PRIOR APPLICATION NUMBER:  09/630,334
;  PRIOR FILING DATE:  2000-07-31
;  PRIOR APPLICATION NUMBER:  09/606,565
;  PRIOR FILING DATE:  2000-06-29
;  PRIOR APPLICATION NUMBER:  09/606,317
;  PRIOR FILING DATE:  2000-06-29
;  PRIOR APPLICATION NUMBER:  09/665,666
;  PRIOR FILING DATE:  2000-09-20
;  PRIOR APPLICATION NUMBER:  09/677,751
;  PRIOR FILING DATE:  2000-09-30
;  NUMBER OF SEQ ID NOS:  162
; SEQ ID NO 125
;   LENGTH: 2715
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-796-753-125

  Query Match             13.9%;  Score 353.4;  DB 11;  Length 2715;
  Best Local Similarity   67.0%;  Pred. No. 3.7e-92;
  Matches  518;  Conservative    0;  Mismatches  251;  Indels    4;  Gaps    1;

Qy        150 CCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGC 209
              || ||||| |||||||    || ||| |||| ||  ||| |||||| | |||||| ||
Db        235 CCAAGGTCAGCTGCTGGTACAAGTACCTCCTTTTCAGCTACAACATCATCTTCTGGTTGG 294

Qy        210 TGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCA 269
              ||||   | |||||  |   |||| || ||||| ||| | || |||||||| || |||
Db        295 CTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCG 354

Qy        270 ACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTG 329
              || || |     |||||    ||   ||  |||||||||||  |||| |  |  |||| |
Db        355 ACCTCACCAAAGTGACCCGGATGCATGGAATCGACCCTGTGGTGCTGGTCCTGATGGTGG 414

Qy        330 GAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTT 389
              | |  || |||| |   ||||| || |||||||| | |||||||| |||||||| | |
Db        415 GCGTGGTGATGTTCACCCTGGGGTTCGCCGGCTGCGTGGGGGCTCTGCGGGAGAATATCT 474

Qy        390 TCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAG 449
              | ||||||| |||||||  |    | | |  ||||||||||||||||||||||     |
Db        475 GCTTGCTCAACTTTTTCTGTGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGG 534

Qy        450 GGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACA 509
              |  |||||||| | ||| |||||||| |  | |||| | ||      ||| |  | |
Db        535 CCGTGCTGGCCTTCCTGTTCCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGA 594

Qy        510 ACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGG 569
              |||| ||||| |||| ||||| || || || || || |||||||| ||||    ||||
Db        595 GCAACATCAAGTCCTACCGGGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGA 654

Qy        570 AATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCA 629
              ||  |      ||||| || ||    || ||| | |||||| |||||||  |||| ||||
Db        655 AAGCTAACCAGTGCTGTGGCGCATATGGCCCTGAAGACTGGGACCTCAACGTCTACTTCA 714

Qy        630 ACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCA 689
              | |||| |    | |  |||||||||   |||||||| |||||||||||||| ||
Db        715 ATTGCAGCGGTGCCAGCTACAGCCGAGAGAAGTGCGGGGTCCCCTTCTCCTGCTGCGTGC 774

Qy        690 GGGACCCTGCG----ATGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGG 745
              || ||||||    | ||  | ||||| |||||||| |||||||| || |  | |||
Db        775 CAGATCCTGCGCAAAAAGTTGTGAACACACAGTGTGGATATGATGTCAGGATTCAGCTGA 834

Qy        746 AGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGC 805
              ||    ||  | | |  |||||   ||| ||||||||  |       | | || |  ||||
Db        835 AGAGCAAGTGGGATGAGTCCATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGC 894

Qy        806 TGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCT 865
              | | |   ||| |   | | |||||||| |||||  | | |||| | ||  | |||||
Db        895 TCCCGCGGAACATTTACATTGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATAT 954

Qy        866 TTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCC 918
              ||||  |||| ||||||    ||  ||  |    |||||| ||||||||||||||
Db        955 TTGGCATCTTCCTGGCAAGGACGCTGATCTCAGACATCGAGGCAGTGAAGGCC 1007

Search completed: November 21, 2003, 21:46:20
Job time : 774 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:42:48 ; Search time 21 Seconds
(without alignments)
469.449 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	412.5	32.2	245	4	US-09-482-273-133	Sequence 133, App
2	265	20.7	60	3	US-09-188-930-181	Sequence 181, App
3	265	20.7	60	3	US-09-188-930-320	Sequence 320, App
4	265	20.7	60	4	US-09-312-283C-181	Sequence 181, App
5	265	20.7	60	4	US-09-312-283C-320	Sequence 320, App
6	240	18.7	253	3	US-09-333-599-4	Sequence 4, Appli
7	240	18.7	253	4	US-09-499-781-4	Sequence 4, Appli
8	239	18.6	219	2	US-08-855-140-4	Sequence 4, Appli
9	236.5	18.4	236	3	US-08-705-771-19	Sequence 19, Appl
10	235.5	18.4	219	2	US-08-855-140-3	Sequence 3, Appli
11	235.5	18.4	219	2	US-08-807-044-3	Sequence 3, Appli
12	235.5	18.4	219	5	PCT-US91-04986-2	Sequence 2, Appli
13	234	18.2	253	3	US-09-333-599-2	Sequence 2, Appli
14	234	18.2	253	4	US-09-499-781-2	Sequence 2, Appli
15	232	18.1	280	2	US-08-855-140-1	Sequence 1, Appli
16	231	18.0	237	3	US-08-808-148-3	Sequence 3, Appli
17	228	17.8	267	3	US-08-430-225A-20	Sequence 20, Appl
18	219.5	17.1	265	2	US-08-807-044-1	Sequence 1, Appli
19	201	15.7	227	1	US-08-254-493-1	Sequence 1, Appli
20	201	15.7	227	2	US-08-253-751-6	Sequence 6, Appli
21	201	15.7	227	2	US-08-453-925-6	Sequence 6, Appli
22	201	15.7	227	4	US-08-403-253A-6	Sequence 6, Appli
23	201	15.7	227	4	US-08-435-816A-6	Sequence 6, Appli
24	201	15.7	228	1	US-08-408-222B-1	Sequence 1, Appli
25	200.5	15.6	281	3	US-08-808-148-4	Sequence 4, Appli
26	190	14.8	252	3	US-08-705-771-17	Sequence 17, Appl
27	163	12.7	241	3	US-08-808-148-1	Sequence 1, Appli
28	163	12.7	241	3	US-09-020-956-114	Sequence 114, App
29	163	12.7	241	3	US-09-030-607-114	Sequence 114, App
30	163	12.7	241	4	US-09-439-313-114	Sequence 114, App
31	163	12.7	241	4	US-09-352-616A-114	Sequence 114, App
32	163	12.7	241	4	US-09-232-149A-114	Sequence 114, App
33	134.5	10.5	204	4	US-09-149-476-429	Sequence 429, App
34	118.5	9.2	260	3	US-08-957-130-15	Sequence 15, Appl
35	108	8.4	258	3	US-08-957-130-13	Sequence 13, Appl
36	96.5	7.5	198	4	US-09-328-352-8198	Sequence 8198, Ap
37	94.5	7.4	101	3	US-08-905-223-443	Sequence 443, App
38	87	6.8	384	3	US-08-852-824-4	Sequence 4, Appli
39	87	6.8	384	4	US-09-731-030A-11	Sequence 11, Appl
40	87	6.8	384	4	US-09-731-030A-13	Sequence 13, Appl
41	82.5	6.4	361	4	US-09-328-352-5459	Sequence 5459, Ap
42	82.5	6.4	397	4	US-09-328-352-7457	Sequence 7457, Ap
43	81	6.3	123	4	US-09-149-476-693	Sequence 693, App
44	80.5	6.3	140	1	US-08-133-347-5	Sequence 5, Appli
45	80.5	6.3	488	4	US-09-328-352-8050	Sequence 8050, Ap

ALIGNMENTS

RESULT 1
US-09-482-273-133
; Sequence 133, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: PZ030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (245)
; OTHER INFORMATION: Xaa equals stop translation
US-09-482-273-133

Query Match 32.2%; Score 412.5; DB 4; Length 245;
Best Local Similarity 44.2%; Pred. No. 8.7e-38;
Matches 87; Conservative 36; Mismatches 59; Indels 15; Gaps 3;

```
Qy          36 LAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKF 95
               :|:|::|   |   : :: |      :||  | :||| :| :| | ||||:||||  ||:
Db           2 VAVGVYARLMKHAEAALACL----AVDPAILLIVVGVLMFLLTFCGCIGSLRENICLLQT 57

Qy          96 FSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSC 155
               ||: |  :| |:|| ||| ||| |  | ::: ||| :  ||||:||||||||| |: :||
Db          58 FSLCLTAVFLLQLAAGILGFVFSDKARGKVSEIINNAIVHYRDDLDLQNLIDFGQKKFSC 117

Qy         156 CGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDP--AMSSTPSVAMMSGSNWSWSS 213
               ||      ||: |:||||:: ||||||| ||:|||:  |  |: :|.    |   ::  :|
Db         118 CGGISYKDWSQNMYFNCSEDNPSRERCSVPYSCCLPTPDQAVINTMCGQGMQAFDYLEAS 177

Qy         214 RAPYTPKAVWASLRSGC 230
               :  ||         :||
Db         178 KVIYT---------NGC 185
```

```
RESULT 2
US-09-188-930-181
; Sequence 181, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 181
;   LENGTH: 60
;   TYPE: PRT
;   ORGANISM: mouse
US-09-188-930-181

  Query Match             20.7%;  Score 265;  DB 3;  Length 60;
  Best Local Similarity   75.9%;  Pred. No. 3.1e-22;
  Matches   44;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         11 PEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFV 68
              ||| || |||:||||::|| ||  || |||||| |||||||||::||||| ||||||:
Db          3 PEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGGFDPVWLFL 60


RESULT 3
US-09-188-930-320
; Sequence 320, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 320
;   LENGTH: 60
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-188-930-320

  Query Match             20.7%;  Score 265;  DB 3;  Length 60;
  Best Local Similarity   75.9%;  Pred. No. 3.1e-22;
  Matches   44;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         11 PEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFV 68
              ||| || |||:||||::|| ||  || |||||| |||||||||::||||| ||||||:
Db          3 PEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGGFDPVWLFL 60


RESULT 4
US-09-312-283C-181
; Sequence 181, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 181
;   LENGTH: 60
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-312-283C-181

  Query Match             20.7%;  Score 265;  DB 4;  Length 60;
  Best Local Similarity   75.9%;  Pred. No. 3.1e-22;
  Matches   44;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         11 PEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFV 68
              ||| || |||:||||::|| ||  || |||||| |||||||||::||||| ||||||:
Db          3 PEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGGFDPVWLFL 60


RESULT 5
US-09-312-283C-320
; Sequence 320, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 320
;   LENGTH: 60
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-312-283C-320

  Query Match             20.7%;  Score 265;  DB 4;  Length 60;
  Best Local Similarity   75.9%;  Pred. No. 3.1e-22;
  Matches   44;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         11 PEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFV 68
              ||| || |||:||||::|| ||  || |||||| |||||||||::||||| ||||||:
Db          3 PEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGGFDPVWLFL 60


RESULT 6
US-09-333-599-4
; Sequence 4, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
;  APPLICANT: Testa, Jacqueline E.
;  APPLICANT: Quigley, James P.
;  APPLICANT: Seandel, Marco
;  TITLE OF INVENTION: MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE: SUNY
;  CURRENT APPLICATION NUMBER: US/09/333,599
;  CURRENT FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 4
;   LENGTH: 253
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-333-599-4

  Query Match             18.7%;  Score 240;  DB 3;  Length 253;
  Best Local Similarity   31.4%;  Pred. No. 1.4e-18;
  Matches   59;  Conservative   32;  Mismatches   85;  Indels   12;  Gaps    6;

Qy          8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              | | :  |    | || || :|  ||: |   :|:|:|    |     || |     |
Db          4 FNEKKATCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASSTYLATA 61

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
              :: || | |: | |  ||    :|   ||: : : | :|| ||:  ||||:|::  :  :
Db         62 YILVVAGVVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGILAYVYQQLLNTE 121

Qy        125 LNFFINNN-VKAYRDD--IDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRER 181
              |   : :  || |       : : :|  |: : |||:    ||  : :    : : ||
Db        122 LKENLKDTMVKRYHQSGHEGVSSAVDKLQQEFHCCGSNNSQDWQDSEWIRSGEAD-SRV- 179

Qy        182 CGVPFSCC 189
                || |||
Db        180 --VPDSCC 185


RESULT 7
US-09-499-781-4
; Sequence 4, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
;  APPLICANT:  Testa, Jacqueline E.
;  APPLICANT:  Quigley, James P.
;  APPLICANT:  Seandel, Marco
;  TITLE OF INVENTION:  MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION:  ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE:  SUNY
;  CURRENT APPLICATION NUMBER: US/09/499,781
;  CURRENT FILING DATE:  2000-02-08
;  PRIOR APPLICATION NUMBER:  09/333,599
;  PRIOR FILING DATE:  1999-06-15
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 253
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-499-781-4

  Query Match             18.7%;  Score 240;  DB 4;  Length 253;
  Best Local Similarity   31.4%;  Pred. No. 1.4e-18;
  Matches   59;  Conservative   32;  Mismatches   85;  Indels   12;  Gaps    6;

Qy          8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              | | :  |    | || || :|  ||: |   :|:|:|    |     || |     |
Db          4 FNEKKATCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASSTYLATA 61

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
              :: || | |: | |  ||    :|   ||: : : | :|| ||:  ||||:|::  :  :
Db         62 YILVVAGVVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGILAYVYQQLLNTE 121

Qy        125 LNFFINNN-VKAYRDD--IDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRER 181
              |   : :  || |       : : :|  |: : |||:    ||  : :    : : ||
Db        122 LKENLKDTMVKRYHQSGHEGVSSAVDKLQQEFHCCGSNNSQDWQDSEWIRSGEAD-SRV- 179

Qy        182 CGVPFSCC 189
                || |||
Db        180 --VPDSCC 185


RESULT 8
US-08-855-140-4
; Sequence 4, Application US/08855140
; Patent No. 5854022
; GENERAL INFORMATION:
;   APPLICANT:  Hillman, Jennifer L.
;   APPLICANT:  Bandman, Olga
;   APPLICANT:  Goli, Surya K.
;   APPLICANT:  Guegler, Karl J.
;   TITLE OF INVENTION:  NEW TRANSMEMBRANE PROTEIN
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;     STREET:  3174 Porter Drive
;     CITY:  Palo Alto
;     STATE:  CA
;     COUNTRY:  USA
;     ZIP:  94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/855,140
;     FILING DATE:  Herewith
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Billings, Lucy J.
;     REGISTRATION NUMBER:  36,749
;     REFERENCE/DOCKET NUMBER:  PF-0296 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415-855-0555
;     TELEFAX:  415-845-4166
;     TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  219 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   IMMEDIATE SOURCE:
;     LIBRARY:  GenBank
;     CLONE:  1279546
US-08-855-140-4

  Query Match             18.6%;  Score 239;  DB 2;  Length 219;
  Best Local Similarity   30.9%;  Pred. No. 1.4e-18;
  Matches   63;  Conservative   29;  Mismatches   72;  Indels   40;  Gaps    7;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
              || || ||::||| |   |  |::      |||  |:  || ||       : |:|| ::
Db         10 KYVLFIFNLLFWVCGCCILGFGIYFLVQNTYGVLFRNLPFLT-LGN-----ILVIVGSII 63

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
              |: | ||:|:::||  ||  | | | :|   |:   || ||::  :   :   :|::::
Db         64 MVVAFLGCMGSIKENKCLLMSFFVLLLIILLAEVTIAILLFVYEQKLNTLVAEGLNDSIQ 123

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPA 194
              |  |         || |     |||  | :||                 | | ||
Db        124 HYHSDNSTMKAWDFIQTQLQCCGVNGSSDWT----------------SGPPSSC------ 161

Qy        195 MSSTPSVAMMSG-----SNWSWSS 213
                || | : |      :|  |:
Db        162 ----PSGADVQGCYNKAKSWFHSN 181


RESULT 9
US-08-705-771-19
; Sequence 19, Application US/08705771
; Patent No. 6054289
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Paul Moore, Reiner Gentz, Hongjin Ji,
;    APPLICANT:  Jian Ni and Jing-Shan Hu
;    TITLE OF INVENTION:  Human Genes, Sequences and
;    TITLE OF INVENTION:  Expression Products
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/705,771
;      FILING DATE:  August 30, 1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MULLINS, J.G.
;      REGISTRATION NUMBER:  33,073
;      REFERENCE/DOCKET NUMBER:  325800-346 (PF196)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  973-994-1700
;      TELEFAX:  973-994-1744
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  236 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-705-771-19

  Query Match             18.4%;  Score 236.5;  DB 3;  Length 236;
  Best Local Similarity   28.9%;  Pred. No. 3e-18;
  Matches   56;  Conservative   34;  Mismatches   83;  Indels   21;  Gaps    6;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVW---LFVVVGGVM 74
              || :| ||::||: |   | :|:|   ::|  : :|:        :|   | :: |  :
Db         10 KYLMFAFNLLFWLGGCGVLGVGIWLAAQQGSFATLSS-----SFPSLWAANLLIITGAFV 64

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
                :|| ||:||::||  ||  | : | |:| ||    || | : | |       :   :
Db         65 MAIGFVGCLGAIKENKCLLLTFFLLLLLVFLLEGTIAILFFAYTDKIDRYAQQDLKKGLH 124

Qy        135 AY--RDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRD 192
               |  : :: | |     |  : |||     ||   :|     | :|    || |||:
Db        125 LYGTQGNVGLTNAWSIIQTDFRCCGVSNYTDW-FEVY------NATR----VPDSCCLEF 173

Qy        193 PAMSSTPSVAMMSG 206
                    : |:: |
Db        174 SESCGLHAPALVEG 187


RESULT 10
US-08-855-140-3
; Sequence 3, Application US/08855140
; Patent No. 5854022
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Guegler, Karl J.
;    TITLE OF INVENTION:  NEW TRANSMEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,140
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0296 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  219 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  180141
US-08-855-140-3

  Query Match             18.4%;  Score 235.5;  DB 2;  Length 219;
  Best Local Similarity   31.0%;  Pred. No. 3.5e-18;
  Matches   54;  Conservative   31;  Mismatches   64;  Indels   25;  Gaps    5;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
              || || ||::||: |   | |::       |||  |: :|| ||      :||:|| ::
Db         10 KYVLFFFNLLFWICGCCILGFGIYLLIHNNPGVLFHNLPSLT-LGN-----VFVIVGSII 63

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
               |: | ||:|:::||  ||  | : | :|   |:   || ||::  : : :   : :::
Db         64 MVVAFLGCMGSIKENKCLLMSFFILLLIILLAEVTLAILLFVYEQKLNEYVAKGLTDSIH 123

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSC 188
               |  |   :  |  | :  |||  | :||                  | | ||
Db        124 RYHSDNSTKAAWDSIQSFLQCCGINGTSDWT----------------SGPPASC 161


RESULT 11
US-08-807-044-3
; Sequence 3, Application US/08807044
; Patent No. 5863735
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/807,044
;      FILING DATE:  Herewith
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0224 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  219 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  180141
US-08-807-044-3

  Query Match             18.4%;  Score 235.5;  DB 2;  Length 219;
  Best Local Similarity   31.0%;  Pred. No. 3.5e-18;
  Matches   54;  Conservative   31;  Mismatches   64;  Indels   25;  Gaps    5;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
              || || ||::||: |   |  |::     |||  |: :|| ||      :||:|| ::
Db         10 KYVLFFFNLLFWICGCCILGFGIYLLIHNNFGVLFHNLPSLT-LGN-----VFVIVGSII 63

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
              |: | ||:|:::||  ||  | : | :|   |:   || ||::  : : :   : :::
Db         64 MVVAFLGCMGSIKENKCLLMSFFILLLIILLAEVTLAILLFVYEQKLNEYVAKGLTDSIH 123

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSC 188
               |  |   :   |  | :  |||  | :||                 | | ||
Db        124 RYHSDNSTKAAWDSIQSFLQCCGINGTSDWT---------------SGPPASC 161


RESULT 12
PCT-US91-04986-2
; Sequence 2, Application PC/TUS9104986
;  GENERAL INFORMATION:
;    APPLICANT:  Seed, Brian
;    APPLICANT:  Allen, Janet
;    APPLICANT:  Aruffo, Alejandro
;    APPLICANT:  Camerini, David
;    APPLICANT:  Lauffer, Leander
;    APPLICANT:  Oquendo, Carmen
;    APPLICANT:  Simmons, David L.
;    APPLICANT:  Stamenkovic, Ivan
;    APPLICANT:  Stengelin, Siegfried
;    APPLICANT:  Amiot, Martine
;    TITLE OF INVENTION:  Rapid Immunoselection Cloning Method
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee & Associates
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/04986
;      FILING DATE:  19910715
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/553,759
;      FILING DATE:  13-JUL-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/498,809
;      FILING DATE:  23-MAR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/379,076
;      FILING DATE:  13-JUL-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/160,416
;      FILING DATE:  25-FEB-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wall, Margaret M.
;      REGISTRATION NUMBER:  33,462
;      REFERENCE/DOCKET NUMBER:  11-88C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  303-499-8080
;      TELEFAX:  303-499-8089
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  219 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US91-04986-2

  Query Match             18.4%;  Score 235.5;  DB 5;  Length 219;
  Best Local Similarity   31.0%;  Pred. No. 3.5e-18;
  Matches   54;  Conservative   31;  Mismatches   64;  Indels   25;  Gaps    5;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
              || || ||::||: |   |  |::     |||  |: :|| ||      :||:|| ::
Db         10 KYVLFFFNLLFWICGCCILGFGIYLLIHNNFGVLFHNLPSLT-LGN-----VFVIVGSII 63

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
              |: | ||:|:::||  ||  | : | :|   |:   || ||::  : : :   : :::
Db         64 MVVAFLGCMGSIKENKCLLMSFFILLLIILLAEVTLAILLFVYEQKLNEYVAKGLTDSIH 123

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSC 188
               |  |   :   |  | :  |||  | :||                 | | ||
Db        124 RYHSDNSTKAAWDSIQSFLQCCGINGTSDWT---------------SGPPASC 161


RESULT 13
US-09-333-599-2
; Sequence 2, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
;  APPLICANT: Testa, Jacqueline E.
;  APPLICANT: Quigley, James P.
;  APPLICANT: Seandel, Marco
;  TITLE OF INVENTION: MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE: SUNY
;  CURRENT APPLICATION NUMBER: US/09/333,599
;  CURRENT FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 253
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-333-599-2

  Query Match             18.2%;  Score 234;  DB 3;  Length 253;
  Best Local Similarity   31.0%;  Pred. No. 6.3e-18;
  Matches   61;  Conservative   29;  Mismatches   77;  Indels   30;  Gaps    7;

Qy          8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              | | :  |    | || || :|  ||: |   :|:|:|    |     || |      |
Db          4 FNEKKTTCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASGTYLATA 61
```

```
Qy          65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               :: || | |: | |  ||    :|   ||: : : | :|| ||:  ||||: :      |
Db          62 YILVVAGTVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGILAYAY----YQQ 117

Qy         125 LNFFINNNVKAYRDDID----------LQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTD 174
               ||  :  |:|   | :           : : :|  |: : |||:    ||  : :
Db         118 LNTELKENLK---DTMTKRYHQPGHEAVTSAVDQLQQEFHCCGSNNSQDWRDSEWIR--- 171

Qy         175 LNPSRERCG--VPFSCC 189
                 |:|  |  || |||
Db         172 ---SQEAGGRVVPDSCC 185


RESULT 14
US-09-499-781-2
; Sequence 2, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
;  APPLICANT:  Testa, Jacqueline E.
;  APPLICANT:  Quigley, James P.
;  APPLICANT:  Seandel, Marco
;  TITLE OF INVENTION:  MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION:  ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE:  SUNY
;  CURRENT APPLICATION NUMBER: US/09/499,781
;  CURRENT FILING DATE:  2000-02-08
;  PRIOR APPLICATION NUMBER:  09/333,599
;  PRIOR FILING DATE:  1999-06-15
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 253
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-499-781-2

  Query Match             18.2%;  Score 234;  DB 4;  Length 253;
  Best Local Similarity   31.0%;  Pred. No. 6.3e-18;
  Matches   61;  Conservative   29;  Mismatches   77;  Indels   30;  Gaps    7;

Qy           8 FQEPEVGC---CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
               | | :  |   | || || :|  ||: |   :|:|:|    |     || |     |
Db           4 FNEKKTTCGTVCLKYLLFTYNCCFWLAGLAVMAVGIWTLALKS--DYISLLASGTYLATA 61

Qy          65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               :: || | |: | |  ||    :|   ||: : : | :|| ||:  ||||: :      |
Db          62 YILVVAGTVVMVTGVLGCCATFKERRNLLRLYFILLLIIFLLEIIAGILAYAY----YQQ 117

Qy         125 LNFFINNNVKAYRDDID----------LQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTD 174
               ||  :  |:|   | :           : : :|  |: : |||:    ||  : :
Db         118 LNTELKENLK---DTMTKRYHQPGHEAVTSAVDQLQQEFHCCGSNNSQDWRDSEWIR--- 171

Qy         175 LNPSRERCG--VPFSCC 189
                 |:|  |  || |||
Db         172 ---SQEAGGRVVPDSCC 185


RESULT 15
US-08-855-140-1
; Sequence 1, Application US/08855140
; Patent No. 5854022
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Guegler, Karl J.
;    TITLE OF INVENTION:  NEW TRANSMEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,140
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0296 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  280 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  MYOMNOT01
;      CLONE:  779308
US-08-855-140-1

  Query Match             18.1%;  Score 232;  DB 2;  Length 280;
  Best Local Similarity   27.5%;  Pred. No. 1.2e-17;
  Matches   60;  Conservative   30;  Mismatches   92;  Indels   36;  Gaps    6;

Qy          15 CCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVM 74
               || || :| ||::||: |   | :|:|    :|  :  |       |    | : :| ::
Db           7 CCLKYMMFLFNLIFWLCGCGLLGVGIWLSVSQGNFATFS--PSFPSLSAANLVIAIGTIV 64

Qy          75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
               | || ||:||::||  ||  | : | :    ||   || ||: | : :     :   :
Db          65 MVTGFLGCLGAIKENKCLLLSFFIVLLVTLLAELILPILFFVYMDKVNENAKKDLKEGLL 124

Qy         135 AY--RDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRD 192
                |   ::: |:|  :  |   |||    ||          |     || ||:
Db         125 LYHTENNVGLKNAWNIIQAEMRCCGVTDYTDW----------YPVLGENTVPDRCCM-- 171

Qy         193 PAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGC 230
                           |    |  ||  :|   |:||
Db         172 --------------ENSQGCGRNGTTP--LW---RTGC 190


Search completed: November 21, 2003, 13:46:00
Job time : 22 secs
```

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:41:03 ; Search time 35 Seconds
(without alignments)
1717.893 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1066	83.1	263	4	Q96FV3	Q96fv3 homo sapien
2	1048	81.7	270	11	Q91VI6	Q91vi6 mus musculu
3	1048	81.7	270	11	Q9D7W4	Q9d7w4 mus musculu
4	885	69.0	266	11	Q9CYT0	Q9cyt0 mus musculu
5	646.5	50.4	270	11	Q8QZY6	Q8qzy6 mus musculu
6	638.5	49.8	270	4	Q8NG11	Q8ng11 homo sapien
7	635.5	49.5	270	4	Q9H0U1	Q9h0u1 homo sapien
8	554	43.2	253	4	Q9BU34	Q9bu34 homo sapien
9	487.5	38.0	123	4	Q9UKB9	Q9ukb9 homo sapien
10	482.5	37.6	269	5	Q9VMJ6	Q9vmj6 drosophila
11	472.5	36.8	283	11	Q8R3S2	Q8r3s2 mus musculu
12	423	33.0	188	4	Q96S98	Q96s98 homo sapien
13	421.5	32.9	291	5	Q9VGV3	Q9vgv3 drosophila
14	419.5	32.7	308	5	Q22495	Q22495 caenorhabdi
15	412.5	32.2	274	5	Q9NB16	Q9nb16 drosophila
16	411.5	32.1	304	5	Q9W4X6	Q9w4x6 drosophila
17	371.5	29.0	302	5	Q9U1L0	Q9u1l0 drosophila
18	336.5	26.2	206	11	Q99L35	Q99l35 mus musculu
19	308.5	24.0	355	4	Q8N548	Q8n548 homo sapien
20	308.5	24.0	355	4	Q9H1Z9	Q9h1z9 homo sapien
21	283	22.1	427	5	O44582	O44582 caenorhabdi
22	283	22.1	451	5	Q95X63	Q95x63 caenorhabdi
23	269	21.0	248	4	Q96SJ8	Q96sj8 homo sapien
24	269	21.0	267	4	Q8WUV1	Q8wuv1 homo sapien
25	265.5	20.7	331	11	Q8VCF5	Q8vcf5 mus musculu
26	259	20.2	349	4	Q9H7Q1	Q9h7q1 homo sapien
27	251	19.6	236	4	Q8NBI9	Q8nbi9 homo sapien
28	242	18.9	225	4	Q8N8E5	Q8n8e5 homo sapien
29	239.5	18.7	245	11	Q99L96	Q99196 mus musculu
30	239	18.6	253	11	Q921J7	Q921j7 mus musculu
31	235.5	18.4	253	11	Q9D1D1	Q9d1d1 mus musculu
32	235	18.3	239	11	Q8BJU2	Q8bju2 mus musculu
33	228.5	17.8	235	11	Q8R3G9	Q8r3g9 mus musculu
34	228.5	17.8	239	5	Q9NB10	Q9nb10 manduca sex
35	228.5	17.8	247	13	Q9PTE0	Q9pte0 gallus gall
36	222.5	17.3	291	5	Q9NB09	Q9nb09 manduca sex
37	222	17.3	224	13	Q9IBC9	Q9ibc9 gallus gall
38	222	17.3	248	5	Q9VLH1	Q9vlh1 drosophila
39	220	17.1	249	11	Q8BT06	Q8bt06 mus musculu
40	219.5	17.1	235	11	O55158	O55158 rattus norv
41	218.5	17.0	267	5	Q9V3E5	Q9v3e5 drosophila
42	214	16.7	227	5	Q8T9S4	Q8t9s4 dermacentor
43	214	16.7	236	5	Q9NB08	Q9nb08 manduca sex
44	210	16.4	236	11	Q91V78	Q91v78 mus musculu
45	209	16.3	245	5	Q9VID1	Q9vid1 drosophila

ALIGNMENTS

RESULT 1
Q96FV3
ID Q96FV3 PRELIMINARY; PRT; 263 AA.
AC Q96FV3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to RIKEN cDNA 2210021G21 gene.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC010405; AAH10405.1; -.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
SQ SEQUENCE 263 AA; 29569 MW; 820189FB903DDB14 CRC64;

Query Match 83.1%; Score 1066; DB 4; Length 263;
Best Local Similarity 99.5%; Pred. No. 5.4e-83;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
```

```
Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 2
Q91VI6
ID   Q91VI6      PRELIMINARY;      PRT;   270 AA.
AC   Q91VI6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   RIKEN cDNA 2210021G21 gene.
GN   2210021G21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC010346; AAH10346.1; -.
DR   MGD; MGI:1921507; 2210021G21Rik.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   270 AA;  30125 MW;  660FE581C1EA42DA CRC64;

  Query Match             81.7%;  Score 1048;  DB 11;  Length 270;
  Best Local Similarity   97.4%;  Pred. No. 1.9e-81;
  Matches  189;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              |||||| ||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQQFQDPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELAAGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDLDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 3
Q9D7W4
ID   Q9D7W4      PRELIMINARY;      PRT;   270 AA.
AC   Q9D7W4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2210021G21Rik protein.
GN   2210021G21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Stomach;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK008761; BAB25880.1; -.
DR   MGD; MGI:1921507; 2210021G21Rik.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   270 AA;  30126 MW;  220FF195C1EA42C1 CRC64;

  Query Match             81.7%;  Score 1048;  DB 11;  Length 270;
  Best Local Similarity   97.4%;  Pred. No. 1.9e-81;
  Matches  189;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              |||||| ||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQQFQDPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELAAGILAFVFKDW 120
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELAAGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDLDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 4
Q9CYT0
ID   Q9CYT0      PRELIMINARY;      PRT;   266 AA.
AC   Q9CYT0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   2810455A09Rik protein.
GN   TM4SF9 OR 2810455A09RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
```

```
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK013350; BAB28804.1; -.
DR   MGD; MGI:1928096; Tm4sf9.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   266 AA;  30040 MW;  A500CFA6192D091D CRC64;

  Query Match             69.0%;  Score 885;  DB 11;  Length 266;
  Best Local Similarity   79.9%;  Pred. No. 1.4e-67;
  Matches  155;  Conservative   16;  Mismatches   21;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| |||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
               ||||||:|||||| :|||||||||||||||||||||||||||:||||||  |:||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::||||||||||||||| |||| |||| | :||||||||||| | |||
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 178

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||| :|||
Db        179 RCGVPFSCCTKDPA 192


RESULT 5
Q8QZY6
ID   Q8QZY6      PRELIMINARY;      PRT;   270 AA.
AC   Q8QZY6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to region containing hypothetical protein MGC11352, slingshot
DE   (Hypothetical protein) (Similar to transmembrane 4 superfamily member
DE   9 homolog).
GN   D14ERTD226E OR AA958793.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Salivary gland;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Pituitary;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; BC024611; AAH24611.1; -.
DR   EMBL; BC025568; AAH25568.1; -.
DR   EMBL; BC026574; AAH26574.1; -.
DR   EMBL; AK030593; BAC27035.1; -.
DR   MGD; MGI:1196325; D14Ertd226e.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   270 AA;  30674 MW;  814DF8AF38E360CB CRC64;

  Query Match             50.4%;  Score 646.5;  DB 11;  Length 270;
  Best Local Similarity   50.6%;  Pred. No. 2.9e-47;
  Matches  122;  Conservative   34;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   || |  || || :|||||: | :|| :||||| ||||||::: :| | |:|||
Db          3 YYRYSNAEVSCWYKYLLFSYNIVFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRLHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
              | ::|| ||  ||||||:||||||  |||||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLKFFCGAIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
                 | :|:|:||||||||||||  |:  |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :   | |               | |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDEFIFTKGCIQALEGWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 6
Q8NG11
ID   Q8NG11      PRELIMINARY;      PRT;   270 AA.
AC   Q8NG11;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   DC-TM4F2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang W., Li N., Wan T., Cao X.;
RT   "Identification of novel membrane proteins.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF311903; AAM94899.1; -.
DR   InterPro; IPR000301; Transmem_4.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   270 AA;  30691 MW;  C74D64CDC52DF107 CRC64;

  Query Match             49.8%;  Score 638.5;  DB 4;  Length 270;
  Best Local Similarity   49.4%;  Pred. No. 1.4e-46;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :||||| ||||||::: :| | |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
              | ::|| ||  ||||||:||||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
                 | :|:|:|||||||||||  |:  |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :   | |               || |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
```

```
Db        243 I 243


RESULT 7
Q9H0U1
ID   Q9H0U1      PRELIMINARY;      PRT;   270 AA.
AC   Q9H0U1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
GN   DKFZP564B1037.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21154917; PubMed=11230166;
RA   Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA   Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA   Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA   Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA   Wambutt R., Korn B., Klein M., Poustka A.;
RT   "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT   Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL   Genome Res. 11:422-435(2001).
DR   EMBL; AL136638; CAB66573.1; -.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
KW   Hypothetical protein.
SQ   SEQUENCE   270 AA;  30696 MW;  4558912DC5381108 CRC64;

  Query Match             49.5%;  Score 635.5;  DB 4;  Length 270;
  Best Local Similarity   49.0%;  Pred. No. 2.5e-46;
  Matches  118;  Conservative   38;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :||||| ||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:|||||  || ||   : ||||||| :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
              |  :|:|:|||||||||:||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNVIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :  | ||             || |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 8
Q9BU34
ID   Q9BU34      PRELIMINARY;      PRT;   253 AA.
AC   Q9BU34;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Similar to transmembrane 4 superfamily member 9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC002920; AAH02920.1; -.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
KW   Transmembrane.
SQ   SEQUENCE   253 AA;  28876 MW;  F83140442CFCAC16 CRC64;

  Query Match             43.2%;  Score 554;  DB 4;  Length 253;
  Best Local Similarity   45.2%;  Pred. No. 2e-39;
  Matches  109;  Conservative   34;  Mismatches   62;  Indels   36;  Gaps    3;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||                ||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFW----------------GVLSDLTKVTRMHGIDPV 45

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:|||||  || ||   : ||||||| :|||:|:||:||:
Db         46 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 105

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
              |  :|:|:|||||||||:||  |:   |||| || ||:||:||||:  : |||:|||
Db        106 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 165

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :  | ||             || |  |    |: :
Db        166 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 225

Qy        226 L 226
              :
Db        226 I 226


RESULT 9
Q9UKB9
ID   Q9UKB9      PRELIMINARY;      PRT;   123 AA.
AC   Q9UKB9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   F-box protein Fbx23 (Fragment).
GN   FBX23.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20003060; PubMed=10531035;
RA   Cenciarelli C., Chiaur D.S., Guardavaccaro D., Parks W., Vidal M.,
RA   Pagano M.;
RT   "Identification of a family of human F-box proteins.";
RL   Curr. Biol. 9:1177-1179(1999).
DR   EMBL; AF174603; AAF04524.1; -.
DR   Genew; HGNC:13594; FBXO23.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
FT   NON_TER       1      1
FT   NON_TER     123    123
SQ   SEQUENCE   123 AA;  13531 MW;  0A6CDB2AF450D8EF CRC64;

  Query Match             38.0%;  Score 487.5;  DB 4;  Length 123;
  Best Local Similarity   85.0%;  Pred. No. 4.3e-34;
  Matches   96;  Conservative    1;  Mismatches   11;  Indels    5;  Gaps    2;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 72
```

```
Qy          61 LDPVWLFVVVGGVMSVLGFAGC---IGALRENTFLLKFFSVFLGLIFFLELAT 110
               ||||||  | |     :|   |   ||||||||||||||  ||||||||||||
Db          73 LDPVWL--VCGSWRRHVGAGLCWAAIGALRENTFLLKFFXXFLGLIFFLELAT 123


RESULT 10
Q9VMJ6
ID   Q9VMJ6      PRELIMINARY;      PRT;   269 AA.
AC   Q9VMJ6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG9093 protein.
GN   TSP26A OR CG9093.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003612; AAF52320.1; -.
DR   FlyBase; FBgn0031760; Tsp26A.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   269 AA;  30235 MW;  A05A188B3224A165 CRC64;

  Query Match             37.6%;  Score 482.5;  DB 5;  Length 269;
  Best Local Similarity   46.2%;  Pred. No. 2.6e-33;
  Matches   92;  Conservative   35;  Mismatches   65;  Indels    7;  Gaps    5;

Qy           1 MPGKHQHF--QEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDL 58
                 ||    : |  :  |: || || ||  |:: |:   | |::|:||| |||:  ||: |  :
Db           1 MPAAVRKFRRETSEISCCLKYLLFASNVILWLSALLVLSVGIWAWSEKGMFRNIARLHFI 60

Qy          59 GGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFK 118
               ||| :: :::|||  :||| | :|||||| ||  :::|| ::   |:    :||| |
Db          61 -ALDPAFVLIILGGVTFLLGFMGSVGALRENTCLLGAYAIFLSVLLIAEIGFCAVAFVLK 119

Qy         119 D--WIRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYW-SCCGARGPNDWNLNIYFNCTDL 175
               |  ||:||    :   :: ||:| | |||||: || |  |||  || ||: | ||||: :
Db         120 DKGWIKDQATEGLKAFIRHYREDADQQNLIDWIQEDWLQCCGIDGPKDWDSNNYFNCSSI 179

Qy         176 N-PSRERCGVPFSCCVRDP 193
                ||| |||||||| | |
Db         180 AIGSREACGVPFSCCRRRP 198


RESULT 11
Q8R3S2
ID   Q8R3S2      PRELIMINARY;      PRT;   283 AA.
AC   Q8R3S2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to transmembrane 4 superfamily member (Tetraspan NET-7).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC024685; AAH24685.1; -.
DR   InterPro; IPR000301; Transmem_4.
KW   Transmembrane.
SQ   SEQUENCE   283 AA;  31550 MW;  9AF1515D5B0EF5CF CRC64;

  Query Match             36.8%;  Score 472.5;  DB 11;  Length 283;
  Best Local Similarity   43.5%;  Pred. No. 1.9e-32;
  Matches  100;  Conservative   41;  Mismatches   70;  Indels   19;  Gaps    4;

Qy           3 GKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLD 62
               |    | | |    || || ||::|||:  : :|:|::|   |   : :: |     :|
Db          12 GDEFSFVSPLV----KYLLFFFNMLFWVISMVMVAVGVYARLMKHAEAALACL----AVD 63

Qy          63 PVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIR 122
               | | :||| :| :| | ||||:|||| ||: ||: | ::| |:|| ||| ||| |  |
Db          64 PAILLIVVGVLMFLLTFCGCIGSLRENICLLQTFSLCLTIVFLLQLAAGILGFVFSDKAR 123

Qy         123 DQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERC 182
               :::  ||| :  ||||:|||||||| |: :|||     ||: |:|||:: |||||||
Db         124 GKVSEIINNAIVHYRDDLDLQNLIDFGQKKFSCCGGISYRDWSQNMYFNCSEDNPSRERC 183

Qy         183 GVPFSCCVRDP--AMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGC 230
               ||:|||:  |  |: :|     |   ::  :|:  ||          :||
Db         184 SVPYSCCLPTPNQAVINTMCGQGMQALDYLEASKVIYT---------NGC 224


RESULT 12
Q96S98
ID   Q96S98      PRELIMINARY;      PRT;   188 AA.
AC   Q96S98;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Tetraspan protein SB134.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang W., Li N., Wan T., Cao X.;
```

```
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY037146; AAK67627.1; -.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
SQ   SEQUENCE   188 AA;  21441 MW;  D69DD8412129D029 CRC64;

  Query Match             33.0%;  Score 423;  DB 4;  Length 188;
  Best Local Similarity   55.9%;  Pred. No. 2.1e-28;
  Matches   85;  Conservative    0;  Mismatches    1;  Indels   66;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFW---------------------------- 29

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
                                             ||||||||||||||||||||||||||
Db         30 ---------------------------------FSVFLGLIFFLELATGILAFVFKDW 54

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEY 152
              |||||| ||||||||||||||||||||||||
Db         55 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEY 86


RESULT 13
Q9VGV3
ID   Q9VGV3      PRELIMINARY;      PRT;   291 AA.
AC   Q9VGV3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4591 protein.
GN   TSP86D OR CG4591.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003688; AAF54570.1; -.
DR   FlyBase; FBgn0037848; Tsp86D.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   291 AA;  33213 MW;  E57FC55688B65C3D CRC64;

  Query Match             32.9%;  Score 421.5;  DB 5;  Length 291;
  Best Local Similarity   40.8%;  Pred. No. 4.4e-28;
  Matches   97;  Conservative   39;  Mismatches   71;  Indels   31;  Gaps   10;

Qy          7 HFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG---LDP 63
              ||    |  | || :|  | :||: | | |||||:|: :|       | |  |   ||
Db         23 HFS--YVSSCVKYMIFLLNFLFWLFGGLLLAIGVYAFMDK--------LMDGNGWLRLDT 72

Qy         64 VW--------LFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAF 115
              ::        : :: | ::  : ||||:|||||||:||| :|: | | | ||::  |: |
Db         73 IYDVIFNISLVMIIAGVIVFTVSFAGCLGALRENTWLLKLYSMCLLLFFILEMSLAIICF 132

Qy        116 VFKDWIRDQLNF-FINNNVKAYRDDIDLQNLIDFAQEYWSCCGA--RGPNDWNLNIYFNC 172
              ||  ::   | : | :  : :|||| |||| |||||: ::|||    |  ||: | ||||
Db        133 VFPQYMNSFLEYQFTDKIIHSYRDDSDLQNFIDFAQQEFNCCGLSNAGYQDWSKNEYFNC 192

Qy        173 TDLNPSRERCGVPFSCCVRDPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGC 230
              :  :|| ||||||:|||:    :||   | :| |       | |    | :|    |||
Db        193 S--SPSVERCGVPYSCCINATDISS-GLVNIMCGYGVQVRSVAA-ASKRIWT---SGC 243


RESULT 14
Q22495
ID   Q22495      PRELIMINARY;      PRT;   308 AA.
AC   Q22495;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   T14G10.6 protein.
GN   T14G10.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wild A.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z68880; CAA93092.1; -.
DR   WormPep; T14G10.6; CE06452.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   308 AA;  34658 MW;  E37221DDD2ADE78E CRC64;

  Query Match             32.7%;  Score 419.5;  DB 5;  Length 308;
  Best Local Similarity   42.9%;  Pred. No. 6.9e-28;
  Matches   78;  Conservative   32;  Mismatches   71;  Indels    1;  Gaps    1;

Qy         10 EPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVV 69
              | |: || || :| ||::|::||   |  |:||  ||    |: :     ||| |  ::
Db         31 ESEISCCVKYSVFSFNVIFFLLGFGLLLFGVWAQIEKNTFVNMLSKASKLYLDPTWPLLI 90
```

```
Qy          70 VGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFI 129
               ||  :  ::||:||:|:|||||  | |:|  |||:   | : |: |:  :| : : :   :
Db          91 VGFLTFIIGFSGCVGSLRENTSFLTFYSTLLGLLLIAEFSAGVFAYACRDQLDNYIRNLL 150

Qy         130 NNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLN-PSRERCGVPFSC 188
               |: |  |||| ||| |||  || | |||  | :||: | ||:       | |  ||||||
Db         151 NDVVVGYRDDPDLQLLIDSMQETWMCCGINGADDWDRNTYFSIEAREVASPEAGGVPFSC 210

Qy         189 CV 190
               |:
Db         211 CI 212


RESULT 15
Q9NB16
ID   Q9NB16      PRELIMINARY;      PRT;   274 AA.
AC   Q9NB16;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Tetraspanin 3A.
GN   TSP3A OR EG:BACR7C10.6 OR CG10742.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Todres E.Z., Nardi J.B., Robertson H.M.;
RT   "The tetraspanin superfamily in insects.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF274013; AAF90139.1; -.
DR   FlyBase; FBgn0040334; Tsp3A.
DR   InterPro; IPR000301; Transmem_4.
DR   Pfam; PF00335; transmembrane4; 1.
DR   PRINTS; PR00259; TMFOUR.
SQ   SEQUENCE   274 AA;  31141 MW;  FD03CC5D50B63018 CRC64;

  Query Match             32.2%;  Score 412.5;  DB 5;  Length 274;
  Best Local Similarity   41.3%;  Pred. No. 2.4e-27;
  Matches   97;  Conservative   33;  Mismatches   70;  Indels   35;  Gaps    9;

Qy          13 VGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEK-------------GVLSNISALTDL 58
               |  | || :|  | |||: | | | ||::|: :|               |  ||| :  |
Db           9 VSQCVKYMIFLLNFVFWLFGGLLLGIGVYAFRDKWEDANGSVRLENFYDVFLNISLVMIL 68

Qy          59 GGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFK 118
                | |   :|:|         |:||:|||||||||||||:|: | | | ||:|  |: ||
Db          69 AGTD---IFLV--------SFSGCVGALRENTFLLKFYSMCLLLFFLLEMAIAIVCFVCP 117

Qy         119 DWIRDQL-NFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGA--RGPNDWNLNIYFNCTDL 175
                ::   |   | :  : :|||| |||| ||||: : |||    |  ||: | ||||:
Db         118 QYMNTFLEKQFTHKIIHSYRDDPDLQNFIDFAQQEFKCCGLSNSGYQDWSKNEYFNCS-- 175

Qy         176 NPSRERCGVPFSCCVRDPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGC 230
               :|| |:||||:|||:    :||   | :| |     ::  |   | :|    |||
Db         176 SPSVEKCGVPYSCCINATDISS-GLVNIMCGYGVQ-NAPVPEATKLIWT---SGC 225


Search completed: November 21, 2003, 13:44:54
Job time : 37 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:41:43 ; Search time 20 Seconds
(without alignments)
1120.365 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	852	66.4	264	2	A59261	tetraspan TSPAN-5
2	419.5	32.7	308	2	T24912	hypothetical prote
3	283	22.1	427	2	T32652	hypothetical prote
4	243	18.9	238	2	A59265	tetraspan TSPAN-4
5	240.5	18.7	245	2	A59258	tetraspan TSPAN-6
6	240	18.7	244	1	I39368	T-cell acute lymph
7	239.5	18.7	245	2	A59260	tetraspan TSPAN-6
8	235.5	18.4	219	1	A37243	hemopoietic cell s
9	233.5	18.2	219	1	A39574	leukocyte antigen
10	231	18.0	237	1	A36056	tumor-associated a
11	228	17.8	267	1	A46493	metastasis suppres
12	220	17.1	238	1	S43511	CD63/ME491 antigen
13	217	16.9	266	2	I49561	C33/R2/IA4 - mouse
14	214	16.7	238	1	I38016	melanoma-associate
15	214	16.7	238	1	JC2297	CD63 antigen - rab
16	211	16.4	238	1	A46508	CD63/ME491 antigen
17	210	16.4	236	1	A46472	cell surface prote
18	208	16.2	226	2	I49589	antigen - mouse
19	207	16.1	236	1	A35649	cell surface prote
20	206	16.1	226	1	S39262	CD9 antigen - rat
21	206	16.1	282	2	T21696	hypothetical prote
22	202	15.7	228	1	A42929	CD9 antigen - gree
23	201	15.7	228	1	A40402	CD9 antigen [valid
24	200.5	15.6	281	1	B47629	cell surface glyco
25	195.5	15.2	218	1	A43522	23K integral membr
26	189	14.7	226	1	JX0221	CD9 antigen - bovi
27	188.5	14.7	218	1	A40181	23K integral membr
28	187.5	14.6	253	2	A59264	tetraspan TSPAN-3
29	184.5	14.4	222	2	A59263	tetraspan TSPAN-2
30	176.5	13.8	281	1	A47629	cell surface glyco
31	175.5	13.7	244	2	T13615	hypothetical prote
32	172.5	13.4	321	2	T45053	hypothetical prote
33	163.5	12.7	206	2	T25161	hypothetical prote
34	163	12.7	241	2	A59262	tetraspan TSPAN-1
35	162	12.6	242	2	T15361	hypothetical prote
36	150	11.7	233	2	T15620	hypothetical prote
37	149.5	11.7	223	2	T26763	hypothetical prote
38	142	11.1	210	2	I58391	sarcoma amplified
39	135.5	10.6	247	2	T28890	hypothetical prote
40	135	10.5	359	2	T18667	hypothetical prote
41	128.5	10.0	194	2	T25548	hypothetical prote
42	128	10.0	309	2	T34080	hypothetical prote
43	125	9.7	203	2	T22537	hypothetical prote
44	125	9.7	346	1	S10177	retinal degenerati
45	122	9.5	346	1	S03347	retinal degenerati

ALIGNMENTS

RESULT 1
A59261
tetraspan TSPAN-5 - human
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02-Jun-2000
C;Accession: A59261
R;Todd, S.C.; Doctor, V.S.; Levy, S.
Biochim. Biophys. Acta 1399, 101-104, 1998
A;Title: Sequences and expression of six new members of the tetraspanin/TM4SF family.
A;Reference number: A59258; MUID:98390278; PMID:9714763
A;Accession: A59261
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-264 <TOD>
A;Cross-references: GB:AF053455; NID:g2995864; PIDN:AAC69712.1; PID:g2995865
C;Genetics:
A;Gene: TSPAN-5
C;Superfamily: CD9 antigen

Query Match 66.4%; Score 852; DB 2; Length 264;
Best Local Similarity 77.8%; Pred. No. 6.5e-66;
Matches 151; Conservative 16; Mismatches 21; Indels 6; Gaps 2;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| |||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||    |||||:||||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFL----SVFLGIIFFLELTAGVLAFVFKDW 114

Qy        121 IRDQLNPFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| |||||::|||||||||||||| |||| |||| | :|||||||||||| | |||
Db        115 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 174

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||| :|||
Db        175 RCGVPFSCCTKDPA 188
```

RESULT 2
T24912
hypothetical protein T14G10.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24912
R;Wild, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19954
A;Accession: T24912
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

```
A;Residues: 1-308 <WIL>
A;Cross-references: EMBL:Z68880; PIDN:CAA93092.1; GSPDB:GN00022; CESP:T14G10.6
A;Experimental source: clone T14G10
C;Genetics:
A;Gene: CESP:T14G10.6
A;Map position: 4
A;Introns: 20/3; 53/1; 283/3

  Query Match             32.7%;  Score 419.5;  DB 2;  Length 308;
  Best Local Similarity   42.9%;  Pred. No. 1.1e-28;
  Matches   78;  Conservative   32;  Mismatches   71;  Indels    1;  Gaps    1;

Qy         10 EPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVV 69
              | |: || || :| ||::|::||    |  |:|| ||     |: :       ||| |  ::
Db         31 ESEISCCVKYSVFSFNVIFFLLGFGLLLFGVWAQIEKNTFVNMLSKASKLYLDPTWPLLI 90

Qy         70 VGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFI 129
              || :  ::||:||:|:|||||  | |:|  |||:   | : |: |:  :| : : :   :
Db         91 VGFLTFIIGFSGCVGSLRENTSFLTFYSTLLGLLLIAEFSAGVFAYACRDQLDNYIRNLL 150

Qy        130 NNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLN-PSRERCGVPFSC 188
              |: |  |||| ||| |||  || | |||  | :||: | ||:       | |  ||||||
Db        151 NDVVVGYRDDPDLQLLIDSMQETWMCCGINGADDWDRNTYFSIEAREVASPEAGGVPFSC 210

Qy        189 CV 190
              |:
Db        211 CI 212


RESULT 3
T32652
hypothetical protein F39C12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T32652
R;Chissoe, S.; Sansone, J.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid F39C12.
A;Reference number: Z21206
A;Accession: T32652
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-427 <CHI>
A;Cross-references: EMBL:AF039043; PIDN:AAB94196.1; GSPDB:GN00028; CESP:F39C12.3
A;Experimental source: strain Bristol N2; clone F39C12
C;Genetics:
A;Gene: CESP:F39C12.3
A;Map position: X
A;Introns: 42/3; 104/3; 133/3; 164/3; 213/3; 276/3; 336/3

  Query Match             22.1%;  Score 283;  DB 2;  Length 427;
  Best Local Similarity   33.6%;  Pred. No. 8.1e-17;
  Matches   72;  Conservative   40;  Mismatches   78;  Indels   24;  Gaps    7;

Qy         12 EVGCCGKYFLFGFNIVFWVLGALFLAIGLWA----WGEKGVLSNISALTDLGGLDPVWLF 67
              |:  | |: :|  | : :::|   ||:|::     : |  ::  |         |:|  |
Db         25 EISACLKWLVFLLNSIVFLVGVGILALGVYLFIKDFREVKLVDII--------LNPAILI 76

Qy         68 VVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFV--FKDWIRD-Q 124
              : |  : |: | | :||||:| |||| |:  : | : | :|  :: |   : |
Db         77 SIFGFSICVVSFFGFMGALRDNIFLLKCFAACVFLSYILVVAVTLVFFTLFYTDTTEGLS 136

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCG----ARGPNDWNLNIYFNCTDLNPSRE 180
              |: :   || |  : :|  ::|  ||   |||    |:|  |||::  ||||: ||  |
Db        137 ANWLLLYAVKNYHTNRNLAEIMDALQENLECCGVSSIAQGYRDWNMSYQFNCTNSNPQPE 196

Qy        181 RCGVPFSCC----VRDPAMSSTPSV-AMMSGSNW 209
              :||||||||    : : | || | : || |   |
Db        197 KCGVPFSCCRKSVISEAAGSSNPLLPAMRSLECW 230


RESULT 4
A59265
tetraspan TSPAN-4 - human
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02-Jun-2000
C;Accession: A59265
R;Todd, S.C.; Doctor, V.S.; Levy, S.
Biochim. Biophys. Acta 1399, 101-104, 1998
A;Title: Sequences and expression of six new members of the tetraspanin/TM4SF family.
A;Reference number: A59258; MUID:98390278; PMID:9714763
A;Accession: A59265
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-238 <TOD>
A;Cross-references: GB:AF054841; NID:g2997746; PIDN:AAC69717.1; PID:g2997747
C;Genetics:
A;Gene: TSPAN-4
C;Superfamily: CD9 antigen

  Query Match             18.9%;  Score 243;  DB 2;  Length 238;
  Best Local Similarity   29.7%;  Pred. No. 1.2e-13;
  Matches   62;  Conservative   33;  Mismatches   90;  Indels   24;  Gaps    9;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMSVL 77
              || :| ||::||: |   | :|:|    :|  : :|  :     |    | :: |  :  :
Db         10 KYLMFAFNLLFWLGGCGVLGVGIWLAATQGSFATLS--SSFPSLSAANLLIITGAFVMAI 67

Qy         78 GFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVKAY- 136
              || ||:||::||  ||  | : | |:| ||    || | : | |        :   :  |
Db         68 GFVGCLGAIKENKCLLLTFFLLLLLVFLLEATIAILFFAYTDKIDRYAQQDLKKGLHLYG 127

Qy        137 -RDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAM 195
               : :: | |      |  : |||      ||   :|      | :|    || |||     :
Db        128 TQGNVGLTNAWSIIQTDFRCCGVSNYTDW-FEVY------NATR----VPDSCC-----L 171

Qy        196 SSTPSVAMMSGSNWSWSSRAP-YTPKAVW 223
                : |  : :   | |  :|| |    ||
Db        172 EFSESCGLHAPGTW-W--KAPCYETVKVW 197


RESULT 5
A59258
tetraspan TSPAN-6 - human
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02-Jun-2000
C;Accession: A59258
R;Todd, S.C.; Doctor, V.S.; Levy, S.
Biochim. Biophys. Acta 1399, 101-104, 1998
A;Title: Sequences and expression of six new members of the tetraspanin/TM4SF family.
A;Reference number: A59258; MUID:98390278; PMID:9714763
A;Accession: A59258
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-245 <TOD>
A;Cross-references: GB:AF053453; NID:g2995860; PIDN:AAC69710.1; PID:g2995861
C;Genetics:
A;Gene: TSPAN-6
C;Superfamily: CD9 antigen

  Query Match             18.7%;  Score 240.5;  DB 2;  Length 245;
  Best Local Similarity   31.6%;  Pred. No. 2.1e-13;
  Matches   61;  Conservative   30;  Mismatches   83;  Indels   19;  Gaps    6;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              |     :  |   | | |  |  :  :||: | : ||:|:|    |  | |  :| :
Db          1 MASPSRRLQTKPVITCFKSVLLIYTFIFWITGVILLAVGIWG---KVSLENYFSLLNEKA 57

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
               :  :: :  | |: :||  ||    | : ::|| :::|| |:| :||   |: |||:
Db         58 TNVPFVLIATGTVIILLGTFGCFATCRASAWMLKLYAMFLTLVFLVELVAAIVGFVFRHE 117

Qy        121 IRDQLNFFINNNVKA---YRDDIDLQ-NLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLN 176
              |:   | | ||  ||   |    | : : :|  |    |||     ||         || |
Db        118 IK---NSFKNNYEKALKQYNSTGDYRSHAVDKIQNTLHCCGVTDYRDW--------TDTN 166
```

```
Qy        177 PSRERCGVPFSCC 189
                |: | | |||
Db        167 YYSEK-GFPKSCC 178
```

RESULT 6
I39368
T-cell acute lymphoblastic leukemia associated antigen 1 - human
N;Alternate names: cell surface glycoprotein (clone A15); TALLA-1
C;Species: Homo sapiens (man)
C;Date: 23-Feb-1996 #sequence_revision 23-Aug-1996 #text_change 22-Jun-1999
C;Accession: I39368; I54784
R;Emi, N.; Kitaori, K.; Seto, M.; Ueda, R.; Saito, H.; Takahashi, T.
Immunogenetics 37, 193-198, 1993
A;Title: Isolation of a novel cDNA clone showing marked similarity to ME491/CD63 superfa
A;Reference number: I39368; MUID:93131291; PMID:8420826
A;Accession: I39368
A;Molecule type: mRNA
A;Residues: 1-244 <RES>
A;Cross-references: GB:D10653; NID:g285900; PIDN:BAA01501.1; PID:g285901
A;Experimental source: immature T cell line HPB-ALL
R;Takagi, S.; Fujikawa, K.; Imai, T.; Fukuhara, N.; Fukudome, K.; Minegishi, M.; Tsuchiy
Int. J. Cancer 61, 706-715, 1995
A;Title: Identification of a highly specific surface marker of T-cell acute lymphoblasti
A;Reference number: I54784; MUID:95286314; PMID:7768645
A;Accession: I54784
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-244 <RE2>
A;Cross-references: GB:D29808; NID:g475005; PIDN:BAA06191.1; PID:g475006
C;Genetics:
A;Gene: GDB:MXS1; DXS1692E; A15; TALLA-1
A;Cross-references: GDB:202921
A;Map position: Xq11-Xq11
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; surface antigen; transmembrane protein
F;1-11/Domain: intracellular #status predicted <CY1>
F;12-35/Domain: transmembrane #status predicted <TM1>
F;36-51/Domain: extracellular #status predicted <EX1>
F;52-76/Domain: transmembrane #status predicted <TM2>
F;77-80/Domain: intracellular #status predicted <CY2>
F;81-102/Domain: transmembrane #status predicted <TM3>
F;103-207/Domain: extracellular #status predicted <EX2>
F;208-232/Domain: transmembrane #status predicted <TM4>
F;233-244/Domain: intracellular #status predicted <CY3>
F;49,150,153,172,183/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match 18.7%; Score 240; DB 1; Length 244;
  Best Local Similarity 28.2%; Pred. No. 2.3e-13;
  Matches 50; Conservative 34; Mismatches 81; Indels 12; Gaps 2;

```
Qy         16 CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPVWLFVVVGGVMS 75
              | |  |  :: |||: | : ||:|:|    | |   :|      :  :: :  |  :
Db          9 CLKTLLIIYSFVFWITGVILLAVGVWG---KLTLGTYISLIAENSTNAPYVLIGTGTTIV 65

Qy         76 VLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVKA 135
              | |  ||    | : ::|| :::|| |:|  ||  ||  |||:  |:|       : ::
Db         66 VFGLFGCFATCRGSPWMLKLYAMFLSLVFLAELVAGISGFVFRHEIKDTFLRTYTDAMQT 125

Qy        136 YRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRD 192
              |  : :    :|  |   |||| :   :|: : ||            |:| |||: :
Db        126 YNGNDERSRAVDHVQRSLSCCGVQNYTNWSTSPYF---------LEHGIPPSCCMNE 173
```

RESULT 7
A59260
tetraspan TSPAN-6 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02-Jun-2000
C;Accession: A59260
R;Todd, S.C.; Doctor, V.S.; Levy, S.
Biochim. Biophys. Acta 1399, 101-104, 1998
A;Title: Sequences and expression of six new members of the tetraspanin/TM4SF family.
A;Reference number: A59258; MUID:98390278; PMID:9714763
A;Accession: A59260
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-245 <TOD>
A;Cross-references: GB:AF053454; NID:g2995862; PIDN:AAC69711.1; PID:g2995863
C;Genetics:
A;Gene: Tspan-6
C;Superfamily: CD9 antigen

  Query Match 18.7%; Score 239.5; DB 2; Length 245;
  Best Local Similarity 29.5%; Pred. No. 2.5e-13;
  Matches 56; Conservative 32; Mismatches 89; Indels 13; Gaps 3;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              |    :  |   |  | |  |  :  :||: | : ||:|:|    |  | |  :| :
Db          1 MASPSRRLQTKPVITCLKSVLLIYTFIFWITGVILLAVGIWG---KVSLENYFSLLNEKA 57

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
                :  :: :  | |: :|| ||    | : ::|| :::|| ||| :||   |: |||:
Db         58 TNVPFVLIGTGTVIILLGTFGCFATCRTSAWMLKLYAMFLTLIFLVELVAAIVGFVFRHE 117

Qy        121 IRDQLNFFINNNVKAYRDDIDLQN-LIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSR 179
              |::     | :| |    | ::  :|  |    |||     ||    |:: |
Db        118 IKNSFKSNYENALKEYNSTGDYRSEAVDKIQSTLHCCGVTNYGDWKGTNYYSET------ 171

Qy        180 ERCGVPFSCC 189
                | | |||
Db        172 ---GFPKSCC 178
```

RESULT 8
A37243
hemopoietic cell surface glycoprotein CD53 - human
N;Alternate names: pan-leukocyte surface antigen CD53
C;Species: Homo sapiens (man)
C;Date: 30-Dec-1991 #sequence_revision 09-Aug-1996 #text_change 22-Jun-1999
C;Accession: A37243; A45872
R;Amiot, M.
J. Immunol. 145, 4322-4325, 1990
A;Title: Identification and analysis of cDNA clones encoding CD53. A pan-leukocyte anti
A;Reference number: A37243; MUID:91079522; PMID:2258620
A;Accession: A37243
A;Molecule type: mRNA
A;Residues: 1-219 <AMI>
A;Cross-references: GB:M60871; NID:g180140; PIDN:AAA51951.1; PID:g180141
R;Angelisova, P.; Vlcek, C.; Stefanova, I.; Lipoldova, M.; Horejsi, V.
Immunogenetics 32, 281-285, 1990
A;Title: The human leucocyte surface antigen CD53 is a protein structurally similar to
A;Reference number: A45872; MUID:91055810; PMID:1700763
A;Accession: A45872
A;Molecule type: mRNA
A;Residues: 1-219 <ANG>
A;Cross-references: GB:M37033; NID:g180142; PIDN:AAA35663.1; PID:g180143
C;Genetics:
A;Gene: GDB:CD53; MOX44
A;Cross-references: GDB:127521; OMIM:151525
A;Map position: 1p21-1p13.3
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; transmembrane protein
F;1-10/Domain: intracellular #status predicted <CY1>
F;11-36/Domain: transmembrane #status predicted <TM1>
F;37-54/Domain: extracellular #status predicted <EX1>
F;55-75/Domain: transmembrane #status predicted <TM2>
F;76-80/Domain: intracellular #status predicted <CY2>
F;81-106/Domain: transmembrane #status predicted <TM3>
F;107-181/Domain: extracellular #status predicted <EX2>
F;182-204/Domain: transmembrane #status predicted <TM4>
F;205-219/Domain: intracellular #status predicted <CY3>
F;129,148/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match 18.4%; Score 235.5; DB 1; Length 219;
  Best Local Similarity 31.0%; Pred. No. 5e-13;

```
  Matches   54;  Conservative   31;  Mismatches   64;  Indels   25;  Gaps    5;

Qy          18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
               || || ||::||: |   |  |::       |||  |: :|| ||      :||:|| ::
Db          10 KYVLFFFNLLFWICGCCILGFGIYLLIHNNFGVLFHNLPSLT-LGN-----VFVIVGSII 63

Qy          75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
               |: | ||:|:::||  ||  | : | :|   |:   || ||::  : : :   : :::
Db          64 MVVAFLGCMGSIKENKCLLMSFFILLLIILLAEVTLAILLFVYEQKLNEYVAKGLTDSIH 123

Qy         135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSC 188
               |  |   :   |  |  :  |||  | :||       '           | | ||
Db         124 RYHSDNSTKAAWDSIQSFLQCCGINGTSDWT---------------SGPPASC 161


RESULT 9
A39574
leukocyte antigen OX-44 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Aug-1991 #sequence_revision 09-Aug-1996 #text_change 22-Jun-1999
C;Accession: A39574
R;Bellacosa, A.; Lazo, P.A.; Bear, S.E.; Tsichlis, P.N.
Mol. Cell. Biol. 11, 2864-2872, 1991
A;Title: The rat leukocyte antigen MRC OX-44 is a member of a new family of cell surface
A;Reference number: A39574; MUID:91203909; PMID:2017181
A;Accession: A39574
A;Molecule type: mRNA
A;Residues: 1-219 <BEL>
A;Cross-references: GB:M57276; NID:g205897; PIDN:AAA41775.1; PID:g205898
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; transmembrane protein
F;1-10/Domain: intracellular #status predicted <CY1>
F;11-36/Domain: transmembrane #status predicted <TM1>
F;37-54/Domain: extracellular #status predicted <EX1>
F;55-73/Domain: transmembrane #status predicted <TM2>
F;74-79/Domain: intracellular #status predicted <CY2>
F;80-106/Domain: transmembrane #status predicted <TM3>
F;107-181/Domain: extracellular #status predicted <EX2>
F;182-204/Domain: transmembrane #status predicted <TM4>
F;205-219/Domain: intracellular #status predicted <CY3>
F;119,129,148/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             18.2%;  Score 233.5;  DB 1;  Length 219;
  Best Local Similarity   32.8%;  Pred. No. 7.4e-13;
  Matches   63;  Conservative   24;  Mismatches   70;  Indels   35;  Gaps    6;

Qy          18 KYFLFGFNIVFWVLGALFLAIG--LWAWGEKGVL-SNISALTDLGGLDPVWLFVVVGGVM 74
               || || || :||| |   |  |       |:|  |:  || ||      : |:|| ::
Db          10 KYVLFFFNFLFWVCGCCILGFGIHLLVQNTYGILFRNLPFLT-LGN-----VLVIVGSII 63

Qy          75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
               |: | ||:|:::||  ||  | | | ||   |:   || ||::  |   :   :|::::
Db          64 MVVAFLGCMGSIKENKCLLMSFFVLLLLILLAEVTLAILLFVYEKKINTLVAEGLNDSIQ 123

Qy         135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPA 194
               |  |   :   || |   |||  | :||           | | ||
Db         124 RYHSDNSTRMAWDFIQSQLQCCGVNGSSDW---------------ISGPPSSC------ 161

Qy         195 MSSTPSVAMMSG 206
                  || | : |
Db         162 ----PSGADVQG 169


RESULT 10
A36056
tumor-associated antigen CO-029 - human
C;Species: Homo sapiens (man)
C;Date: 16-Nov-1990 #sequence_revision 09-Aug-1996 #text_change 20-Apr-2000
C;Accession: A36056
R;Szala, S.; Kasai, Y.; Steplewski, Z.; Rodeck, U.; Koprowski, H.; Linnenbach, A.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6833-6837, 1990
A;Title: Molecular cloning of cDNA for the human tumor-associated antigen CO-029 and ide
A;Reference number: A36056; MUID:90370878; PMID:2395876
A;Accession: A36056
A;Molecule type: mRNA
A;Residues: 1-237 <SZA>
A;Cross-references: GB:M35252; NID:g180925; PIDN:AAA35709.1; PID:g180926
C;Genetics:
A;Gene: GDB:TM4SF3
A;Cross-references: GDB:9113496; OMIM:600769
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; transmembrane protein
F;2-11/Domain: intracellular #status predicted <CY1>
F;12-33/Domain: transmembrane #status predicted <TM1>
F;34-52/Domain: extracellular #status predicted <EX1>
F;53-78/Domain: transmembrane #status predicted <TM2>
F;79-82/Domain: intracellular #status predicted <CY2>
F;83-107/Domain: transmembrane #status predicted <TM3>
F;108-205/Domain: extracellular #status predicted <EX2>
F;206-232/Domain: transmembrane #status predicted <TM4>
F;233-237/Domain: intracellular #status predicted <CY3>
F;37,118/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             18.0%;  Score 231;  DB 1;  Length 237;
  Best Local Similarity   33.1%;  Pred. No. 1.3e-12;
  Matches   56;  Conservative   24;  Mismatches   67;  Indels   22;  Gaps    4;

Qy          13 VGCCGKYFLFGFNIVFWVLGALFLAIGLWA---------WGEKGVLSNISALTDLGGLDP 63
               |  | || :| || :||: | | ||: :|          :| : | |:      |
Db           4 VSACIKYSMFTFNFLFWLCGILILALAIWVRVSNDSQAIFGSEDVGSSSYVAVD------ 57

Qy          64 VWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRD 123
               : : || :: :||| || ||::|:  :|  | : | ||  |::|||||  |||
Db          58 --ILIAVGAIIMILGFLGCCGAIKESRCMLLLFFIGLLLILLLQVATGILGAVFKSKSDR 115

Qy         124 QLNFFINNNVKAY----RDDIDLQNLIDFAQEYWSCCG-ARGPNDWNLN 167
                :|  :  | |         :   |  |   || : |||   |  ||  |
Db         116 IVNETLYENTKLLSATGESEKQFQEAIIVFQEEFKCCGLVNGAADWGNN 164


RESULT 11
A46493
metastasis suppressor KAI1 - human
N;Alternate names: cell surface glycoprotein KAI1; membrane protein R2, inducible; type
C;Species: Homo sapiens (man)
C;Date: 18-Jun-1993 #sequence_revision 09-Aug-1996 #text_change 22-Jun-1999
C;Accession: I38942; S16156; A46493
R;Dong, J.T.; Lamb, P.W.; Rinker-Schaeffer, C.W.; Vukanovic, J.; Ichikawa, T.; Isaacs,
Science 268, 884-886, 1995
A;Title: KAI1, a metastasis suppressor gene for prostate cancer on human chromosome 11p
A;Reference number: I38942; MUID:95273964; PMID:7754374
A;Accession: I38942
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U20770; NID:g806805; PIDN:AAC50133.1; PID:g806806
R;Gaugitsch, H.W.; Hofer, E.; Huber, N.E.; Schnabl, E.; Baumruker, T.
Eur. J. Immunol. 21, 377-383, 1991
A;Title: A new superfamily of lymphoid and melanoma cell proteins with extensive homolo
A;Reference number: S16156; MUID:91153380; PMID:1842498
A;Accession: S16156
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-267 <GAU>
A;Cross-references: EMBL:X53795; NID:g35832; PIDN:CAA37804.1; PID:g35833
A;Note: the authors translated the codon AGC for residue 50 as Thr
R;Imai, T.; Fukudome, K.; Takagi, S.; Nagira, M.; Furuse, M.; Fukuhara, N.; Nishimura,
J. Immunol. 149, 2879-2886, 1992
A;Title: C33 antigen recognized by monoclonal antibodies inhibitory to human T cell leu
D9, CD37, CD53, and CD63.
A;Reference number: A46493; MUID:93017900; PMID:1401919
A;Accession: A46493
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-239,'MV',242-267 <IMA>
A;Cross-references: GB:S48196; NID:g258294; PIDN:AAB23825.1; PID:g258295
```

```
A;Experimental source: T-cell line MOLT-4
A;Note: sequence extracted from NCBI backbone (NCBIP:117149)
C;Genetics:
A;Gene: GDB:KAI1
A;Cross-references: GDB:134216; OMIM:600623
A;Map position: 11p11.2-11p11.2
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; transmembrane protein
F;1-10/Domain: intracellular #status predicted <CY1>
F;11-36/Domain: transmembrane #status predicted <TM1>
F;37-57/Domain: extracellular #status predicted <EX1>
F;58-78/Domain: transmembrane #status predicted <TM2>
F;79-83/Domain: intracellular #status predicted <CY2>
F;84-108/Domain: transmembrane #status predicted <TM3>
F;109-227/Domain: extracellular #status predicted <EX2>
F;228-252/Domain: transmembrane #status predicted <TM4>
F;253-265/Domain: intracellular #status predicted <CY3>
F;129,157,198/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             17.8%;  Score 228;  DB 1;  Length 267;
  Best Local Similarity   30.0%;  Pred. No. 2.7e-12;
  Matches   68;  Conservative   37;  Mismatches   88;  Indels   34;  Gaps   11;

Qy         18 KYFLFGFNIVFWVLGALFLAIGLWAWGEK----GVLSNISALTDLGGLDPVWLFVVVGGV 73
              ||||| ||::|::|||: |  |:|   :|     ||   |:   :|      ::|: || |
Db         10 KYFLFLFNLIFFILGAVILGFGVWILADKSSFISVLQTSSSSLRMG----AYVFIGVGAV 65

Qy         74 MSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNV 133
                ::|| |||||: |   ||  :  || ||   ::  | | :     :: ::   :   :
Db         66 TMLMGFLGCIGAVNEVRCLLGLYFAFLLLILIAQVTAGALFYFNMGKLKQEMGGIVTELI 125

Qy        134 KAY---RDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTD----LNPSRERCGVPF 186
               : |   |:|  ||:  |: |     |||       |    ::| ||    :|  |      |
Db        126 RDYNSSRED-SLQDAWDYVQAQVKCCG------W--VSFYNWTDNAELMN--RPEVTYPC 174

Qy        187 SCCVR---DPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGC 230
              || |:   | ::|         | | : |   |       |   : ||
Db        175 SCEVKGEEDNSLSVRKGFCEAPG-NRTQSGNHPED----WPVYQEGC 216


RESULT 12
S43511
CD63/ME491 antigen homolog - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 09-Aug-1996 #text_change 16-Jun-2000
C;Accession: S43511
R;Miyamoto, H.; Homma, M.; Hotta, H.
Biochim. Biophys. Acta 1217, 312-316, 1994
A;Title: Molecular cloning of the murine homologue of CD63/ME491 and detection of its st
A;Reference number: S43511; MUID:94198294; PMID:8148377
A;Accession: S43511
A;Molecule type: mRNA
A;Residues: 1-238 <MIY>
A;Cross-references: EMBL:D16432; NID:g484052; PIDN:BAA03904.1; PID:g976238
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; lysosome; surface antigen; transmembrane protein
F;1-11/Domain: intracellular #status predicted <CY1>
F;12-35/Domain: transmembrane #status predicted <TM1>
F;36-51/Domain: extracellular #status predicted <EX1>
F;52-76/Domain: transmembrane #status predicted <TM2>
F;77-80/Domain: intracellular #status predicted <CY2>
F;81-103/Domain: transmembrane #status predicted <TM3>
F;104-202/Domain: extracellular #status predicted <EX2>
F;203-228/Domain: transmembrane #status predicted <TM4>
F;229-238/Domain: intracellular #status predicted <CY3>
F;116,130,150,172/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             17.1%;  Score 220;  DB 1;  Length 238;
  Best Local Similarity   31.2%;  Pred. No. 1.2e-11;
  Matches   55;  Conservative   31;  Mismatches   76;  Indels   14;  Gaps    5;

Qy         16 CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSN-ISALTDLGGLDPVWLFVVVGGVM 74
              | |: |:   : |      :|||:     : ||   |:  |  | | || : : ||  :
Db          9 CVKFLLYVLLLAFCACAVGLIAIGV---AVQVVLKQAITHETTAGSLLPV-VIIAVGAFL 64

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
              :: | || || :||  |:  |::|| ||  :|:|  |  :||:| :: : |       ::
Db         65 FLVAFVGCCGACKENYCLMITFAIFLSLIMLVEVAVAIAGYVFRDQVKSEFNKSFQQQMQ 124

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCV 190
               | |      ::|  |:  :|||     ||  ||         |   :  || |||:
Db        125 NYLKDNKTATILDKLQKENNCCGASNYTDWE-NI--------PGMAKDRVPDSCCI 171


RESULT 13
I49561
C33/R2/IA4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Aug-1999
C;Accession: I49561
R;Nagira, M.; Imai, T.; Ishikawa, I.; Uwabe, K.I.; Yoshie, O.
Cell. Immunol. 157, 144-157, 1994
A;Title: Mouse homologue of C33 antigen (CD82), a member of the transmembrane 4 superf
A;Reference number: I49561; MUID:94313678; PMID:8039242
A;Accession: I49561
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-266 <RES>
A;Cross-references: GB:D14883; NID:g984036; PIDN:BAA03602.1; PID:g498368
C;Superfamily: CD9 antigen

  Query Match             16.9%;  Score 217;  DB 2;  Length 266;
  Best Local Similarity   31.5%;  Pred. No. 2.4e-11;
  Matches   51;  Conservative   32;  Mismatches   67;  Indels   12;  Gaps    4;

Qy         14 GC--CGKYFLFGFNIVFWVLGALFLAIGLWAWGEK----GVLSNISALTDLGGLDPVWLF 67
              ||    ||||| ||::|::|||: |  |:|   :|      ||   |:   :|      ::|
Db          4 GCVKVTKYFLFLFNLLFFILGAVILGFGVWILADKNSFISVLQTSSSSLQVG----AYVF 59

Qy         68 VVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNF 127
              : || :  |:|| |||||: |   ||  : ||| ||   ::  |:| :   | :: ::
Db         60 IGVGAITIVMGFLGCIGAVNEVRCLLGLYFVFLLLILIAQVTVGVLFYFNADKLKKEMGN 119

Qy        128 FINNNVKAYRDDI--DLQNLIDFAQEYWSCCGARGPNDWNLN 167
              : : ::  |  :     :   |: |    |||     :|  |
Db        120 TVMDIIRNYTANATSSREEAWDYVQAQVKCCGWVSHYNWTEN 161


RESULT 14
I38016
melanoma-associated antigen CD63 [validated] - human
N;Alternate names: antigen ME491; lysosomal membrane glycoprotein CD63; ME491/CD63 ant
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 09-Aug-1996 #text_change 08-Dec-2000
C;Accession: I38016; S01418; A39514; B35826; A61177; A61173; A56782
R;Hotta, H.; Miyamoto, H.; Hara, I.; Takahashi, N.; Homma, M.
Biochem. Biophys. Res. Commun. 185, 436-442, 1992
A;Title: Genomic structure of the ME491/CD63 antigen gene and functional analysis of t
A;Reference number: I38016; MUID:92287132; PMID:1599482
A;Accession: I38016
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-238 <RES>
A;Cross-references: EMBL:X62654; NID:g430755; PIDN:CAA44519.1; PID:g430756
R;Hotta, H.; Ross, A.H.; Huebner, K.; Isobe, M.; Wendeborn, S.; Chao, M.V.; Ricciardi,
Cancer Res. 48, 2955-2962, 1988
A;Title: Molecular cloning and characterization of an antigen associated with early st
A;Reference number: S01418; MUID:88210273; PMID:3365686
A;Accession: S01418
A;Molecule type: mRNA
A;Residues: 1-238 <HOT>
A;Cross-references: EMBL:X07982; NID:g34526; PIDN:CAA30792.1; PID:g34527
R;Metzelaar, M.J.; Wijngaard, P.L.J.; Peters, P.J.; Sixma, J.J.; Nieuwenhuis, H.K.; Cl
J. Biol. Chem. 266, 3239-3245, 1991
A;Title: CD63 antigen. A novel lysosomal membrane glycoprotein, cloned by a screening
A;Reference number: A39514; MUID:91131632; PMID:1993697
```

```
A;Accession: A39514
A;Molecule type: mRNA
A;Residues: 1-238 <MET>
A;Cross-references: GB:M58485
R;Rapp, G.; Freudenstein, J.; Klaudiny, J.; Mucha, J.; Wempe, F.; Zimmer, M.; Scheit, K.
DNA Cell Biol. 9, 479-485, 1990
A;Title: Characterization of three abundant mRNAs from human ovarian granulosa cells.
A;Reference number: A35826; MUID:91025550; PMID:2171551
A;Accession: B35826
A;Molecule type: mRNA
A;Residues: 1-238 <RAP>
A;Cross-references: GB:M59907; NID:g189383; PIDN:AAA63235.1; PID:g189384
A;Note: the authors did not translate the codons for residues 205 through 224
R;Azorsa, D.O.; Hyman, J.A.; Hildreth, J.E.K.
Blood 78, 280-284, 1991
A;Title: CD63/Pltgp40: a platelet activation antigen identical to the stage-specific, me
A;Reference number: A61177; MUID:91300080; PMID:2070066
A;Accession: A61177
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2-68,'P',70-238 <AZO>
R;Hildreth, J.E.K.; Derr, D.; Azorsa, D.O.
Blood 77, 121-132, 1991
A;Title: Characterization of a novel self-associating Mr 40,000 platelet glycoprotein.
A;Reference number: A61173; MUID:91084576; PMID:1984792
A;Accession: A61173
A;Molecule type: protein
A;Residues: 2-8,'X',10-16,'XX',19-21 <HIL>
R;Wang, M.X.; Earley Jr., J.J.; Shields, J.A.; Donoso, L.A.
Arch. Ophthalmol. 110, 399-404, 1992
A;Title: An ocular melanoma-associated antigen. Molecular characterization.
A;Reference number: A56782; MUID:92181348; PMID:1339263
A;Accession: A56782
A;Molecule type: mRNA
A;Residues: 1-238 <WAN>
A;Cross-references: GB:S93788; NID:g246538; PIDN:AAB21617.1; PID:g246539
A;Experimental source: uveal melanoma
A;Note: sequence extracted from NCBI backbone (NCBIN:93788, NCBIP:93790)
C;Genetics:
A;Gene: GDB:CD63; MLA1
A;Cross-references: GDB:120186; OMIM:155740
A;Map position: 12q12-12q13
A;Introns: 22/3; 85/3; 110/3; 142/3; 189/3; 217/3
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; lysosome; surface antigen; transmembrane protein
F;2-238/Product: melanoma-associated antigen ME491 #status experimental <MAT>
F;2-11/Domain: intracellular #status predicted <CY1>
F;12-35/Domain: transmembrane #status predicted <TM1>
F;36-51/Domain: extracellular #status predicted <EX1>
F;52-76/Domain: transmembrane #status predicted <TM2>
F;77-80/Domain: intracellular #status predicted <CY2>
F;81-103/Domain: transmembrane #status predicted <TM3>
F;104-202/Domain: extracellular #status predicted <EX2>
F;203-228/Domain: transmembrane #status predicted <TM4>
F;229-238/Domain: intracellular #status predicted <CY3>
F;130,150,172/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             16.7%;  Score 214;  DB 1;  Length 238;
  Best Local Similarity   28.1%;  Pred. No. 3.8e-11;
  Matches   61;  Conservative   37;  Mismatches   83;  Indels   36;  Gaps    8;

Qy         16 CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSN--ISALTDLGGLDPVWLFVVVGGV 73
              | |: |:   : |       :|:|:   | : |||   |   |  | | || : : ||
Db          9 CVKFLLYVLLLAFCACAVGLIAVGV---GAQLVLSQTIIQGATP-GSLLPV-VIIAVGVF 63

Qy         74 MSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNV 133
              : :: | || || :||  |:  |::|| ||  :|:|  |  :||:| :  : |      :
Db         64 LFLVAFVGCCGACKENYCLMITFAIFLSLIMLVEVAAAIAGYVFRDKVMSEFNNNFRQQM 123

Qy        134 KAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDP 193
              : |  :    :::|  |  : ||||    ||         :  ||  :  || |||:
Db        124 ENYPKNNHTASILDRMQADFKCCGAANYTDW---------EKIPSMSKNRVPDSCCI--- 171

Qy        194 AMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGC 230
                 :| :  | |       :  ||:    : ||
Db        172 ------NVTVGCGIN--------FNEKAIH---KEGC 191


RESULT 15
JC2297
CD63 antigen - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 28-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 16-Jun-2000
C;Accession: JC2297
R;Sohma, Y.; Suzuki, T.; Sasano, H.; Nagura, H.; Nose, M.; Yamamoto, T.
Cell Struct. Funct. 19, 219-225, 1994
A;Title: Increased mRNA for CD63 antigen in atherosclerotic lesions of Watanable herit
A;Reference number: JC2297; MUID:95120837; PMID:7820873
A;Accession: JC2297
A;Molecule type: mRNA
A;Residues: 1-238 <SOH>
A;Cross-references: DDBJ:D21264; NID:g684973; PIDN:BAA04804.1; PID:g684974
A;Experimental source: aorta
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; lysosome; surface antigen; transmembrane protein
F;1-11/Domain: intracellular #status predicted <CY1>
F;12-35/Domain: transmembrane #status predicted <TM1>
F;36-51/Domain: extracellular #status predicted <EX1>
F;52-76/Domain: transmembrane #status predicted <TM2>
F;77-80/Domain: intracellular #status predicted <CY2>
F;81-103/Domain: transmembrane #status predicted <TM3>
F;104-202/Domain: extracellular #status predicted <EX2>
F;203-228/Domain: transmembrane #status predicted <TM4>
F;229-238/Domain: intracellular #status predicted <CY3>
F;125,130,150,172/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             16.7%;  Score 214;  DB 1;  Length 238;
  Best Local Similarity   31.2%;  Pred. No. 3.8e-11;
  Matches   55;  Conservative   32;  Mismatches   75;  Indels   14;  Gaps    5;

Qy         16 CGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLS-NISALTDLGGLDPVWLFVVVGGVM 74
              | |: |:   : |       :|:|:   | : |||  |:     | | || : : ||  :
Db          9 CVKFLLYVLLLAFCACAVGLIAVGV---GAQLVLSQTITHGATPGSLLPV-VIIAVGAFL 64

Qy         75 SVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQLNFFINNNVK 134
              :: | || |  :||  |:  |::|| ||  :|:|  |  :||:| :  : |      ::
Db         65 FLVAFVGCCGTCKENYCLMITFAIFLSLIMLVEVAAAIAGYVFRDKVMSEFNKDFRQQMQ 124

Qy        135 AYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCV 190
              | |     ::|  |: ::|||    ||         |    :|:|  || ||||
Db        125 NYSTDNQTALILDRMQKDFTCCGAANYTDW-------ATIPGMTRDR--VPDSCCV 171


Search completed: November 21, 2003, 13:45:26
Job time : 21 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:34:37 ; Search time 40 Seconds
(without alignments)
924.582 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1283	100.0	233	23	AAE13218	Human gene 1 encod
2	1066	83.1	270	22	AAB49504	Clone HNTMH27. Ho
3	1066	83.1	270	23	AAB47950	Human 23228. Homo
4	1066	83.1	329	21	AAB18969	Amino acid sequenc
5	1055	82.2	329	23	ABB90429	Human polypeptide
6	972.5	75.8	330	23	ABP69599	Human polypeptide
7	885	69.0	271	21	AAB53438	Human colon cancer
8	791	61.7	268	22	AAB49505	Clone HE8EJ16 #1.
9	699	54.5	193	20	AAY42381	Amino acid sequenc
10	699	54.5	193	22	AAU39073	Human secreted pro
11	699	54.5	193	23	ABB55782	Human polypeptide
12	640.5	49.9	512	22	ABP98695	Rat gamma-hydroxyb
13	638.5	49.8	270	21	AAY76266	Human secreted pro
14	638.5	49.8	270	22	AAB87034	Human secreted pro
15	638.5	49.8	270	23	AAE18535	Human tetraspan pr
16	637.5	49.7	270	22	AAB87134	Human secreted pro
17	635.5	49.5	270	22	AAB87135	Human secreted pro
18	635.5	49.5	292	22	ABG27829	Novel human diagno
19	634.5	49.5	270	22	AAB87136	Human secreted pro
20	634.5	49.5	270	22	AAB87137	Human secreted pro
21	625.5	48.8	270	21	AAY76133	Human secreted pro
22	619.5	48.3	270	21	AAB23036	Human tetraspanin-
23	614.5	47.9	253	22	AAB87039	Human TANGO 339 pe
24	556	43.3	200	22	ABG16480	Novel human diagno
25	515.5	40.2	228	22	AAB87035	Human mature TANGO
26	482.5	37.6	122	21	AAY83082	F-box protein FBP-
27	482.5	37.6	122	23	AAO22468	Human F-box protei
28	482.5	37.6	269	22	ABB64234	Drosophila melanog
29	471.5	36.7	283	22	AAU27652	Human protein AFP2
30	471.5	36.7	283	23	ABP69619	Human polypeptide
31	453.5	35.3	193	22	AAB87037	Human TANGO 339 tr
32	437.5	34.1	368	23	ABB89649	Human polypeptide
33	429	33.4	120	22	AAG75522	Human colon cancer
34	425	33.1	213	21	AAB12120	Hydrophobic domain
35	421.5	32.9	291	22	ABB60781	Drosophila melanog
36	412.5	32.2	244	22	AAE06071	Human gene 31 enco
37	412.5	32.2	244	23	ABG33893	Human secreted pro
38	412.5	32.2	245	21	AAY87094	Human secreted pro
39	411.5	32.1	209	21	AAB23047	Human tetraspanin-
40	411.5	32.1	304	22	ABB65216	Drosophila melanog
41	383	29.9	111	22	AAB49510	Clone HE8EL16 #2.
42	373	29.1	248	22	ABG16481	Novel human diagno
43	331	25.8	104	21	AAY64846	Human 5' EST relat
44	324	25.3	294	19	AAW70319	Secreted protein B
45	324	25.3	294	21	AAY99374	Human PRO1311 (UNQ

ALIGNMENTS

RESULT 1
AAE13218
ID AAE13218 standard; Protein; 233 AA.
XX
AC AAE13218;
XX
DT 12-FEB-2002 (first entry)
XX
DE Human gene 1 encoded TM4SF receptor protein HOFOB55, SEQ ID NO:4.
XX
KW Human; 4-transmembrane superfamily; TM4SF receptor; breast cancer;
KW ovarian cancer; immune disorder; Addison's disease; wound healing;
KW gene therapy; autoimmune haemolytic anaemia; autoimmune thyroiditis;
KW diabetes mellitus; allergy; Crohn's disease; multiple sclerosis;
KW rheumatoid arthritis; ulcerative colitis; cardiovascular disorder;
KW myocardial ischaemia; neurological disease; antifungal; antiviral;
KW antibacterial; cerebral anoxia; epilepsy; infectious disease;
KW antiparasitic; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 158..164
FT /note= "Immunogenic epitope"
FT Region 174..182
FT /note= "Immunogenic epitope"
FT Region 212..218
FT /note= "Immunogenic epitope"
FT Region 228..233
FT /note= "Immunogenic epitope"
XX
PN WO200177173-A1.

XX
PD 18-OCT-2001.
XX
PF 05-APR-2001; 2001WO-US11130.
XX
PR 10-APR-2000; 2000US-195336P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Shi Y, Ruben SM;
XX
DR WPI; 2002-017447/02.
DR N-PSDB; AAD21883.
XX
PT Novel isolated protein, a member of 4-transmembrane superfamily of
PT receptor polypeptides, useful in the prevention, treatment and
PT diagnosis of cancer, immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 11; Page 267-268; 271pp; English.
XX
CC The invention relates to human 4-transmembrane superfamily (TM4SF)
CC receptor polypeptides and polynucleotides. Sequences of the invention
CC are useful for preventing, treating, ameliorating or diagnosing a
CC pathological condition or a susceptibility to a pathological condition.
CC TM4SF polypeptides are useful for screening molecules which modify
CC their activity. TM4SF nucleic acids, protein, antibodies, agonists and
CC antagonists are useful in the diagnosis, treatment and prevention of
CC cancer, particularly breast and ovarian cancer, and other cancers of
CC the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital; immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. TM4SF polynucleotides are
CC also useful in gene therapy. The present sequence is human TM4SF
CC receptor protein.
XX
SQ Sequence 233 AA;

Query Match 100.0%; Score 1283; DB 23; Length 233;
Best Local Similarity 100.0%; Pred. No. 9.4e-128;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGCRTT 233
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RCGVPFSCCVRDPAMSSTPSVAMMSGSNWSWSSRAPYTPKAVWASLRSGCRTT 233
```

RESULT 2
AAB49504
ID AAB49504 standard; Protein; 270 AA.
XX
AC AAB49504;
XX
DT 09-MAR-2001 (first entry)
XX
DE Clone HNTMH27.
XX
KW Human; 4 transmembrane superfamily receptor protein;
KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW bacterial infection; viral; fungal.
XX
OS Homo sapiens.
XX
PN WO200070076-A1.
XX
PD 23-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13504.
XX
PR 19-MAY-1999; 99US-0135122.
PR 03-JUN-1999; 99US-0137797.
PR 11-JUN-1999; 99US-0138573.
PR 18-AUG-1999; 99US-0149447.
PR 28-JAN-2000; 2000US-0178770.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;
PI Rosen CA;
XX
DR WPI; 2001-007502/01.
DR N-PSDB; AAC90014.
XX
PT Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT superfamily receptor protein, useful for diagnosing, treating and/or
PT preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS Claim 11; Pages 280-281; 297pp; English.
XX
CC The present invention relates to isolated nucleic acids and proteins
CC encoding human soluble 4 transmembrane superfamily receptor protein (see
CC AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
CC protein. The present protein can be used to screen for binding partners
CC and molecules which modify its activity. Antibodies specific for the
CC present protein can be used to treat and/or prevent diseases associated
CC with abberant expression or activity of the present protein e.g.
CC endocrine disorders e.g. Addison's disease, (cardio)vascular diseases
CC e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural
CC disorders e.g. Alzheimer's and Parkinson's disease, reproductive
CC disorders, skin disorders e.g. psoriasis, renal system disorders e.g.
CC nephritis, (auto)immune system disorders e.g. graft vs. host disease,
CC hyperproliferative disorders e.g. neoplasms of the pancreas, ocular
CC disorders e.g. glaucoma and infections caused by bacteria, viruses and
CC fungi.
XX
SQ Sequence 270 AA;

Query Match 83.1%; Score 1066; DB 22; Length 270;
Best Local Similarity 99.5%; Pred. No. 1.2e-104;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194
```

RESULT 3

```
AAB47950
ID   AAB47950 standard; Protein; 270 AA.
XX
AC   AAB47950;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human 23228.
XX
KW   Tetraspanin; 23228; cell surface protein; transmembrane domain;
KW   extracellular growth factor; HB-EGF; TGF-alpha; amphiregulin;
KW   diabetes mellitus; arthritis; multiple sclerosis; encephalomyelitis;
KW   dermatitis; Crohn's disease; and asthma; cancer; metastasis;
KW   viral infection; cellular signalling activity; cell proliferation;
KW   cell motility; CD81; B-Cell antigen receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..18
FT                   /label= N-terminal_domain
FT   Domain          18..263
FT                   /label= Tetraspanin_domain
FT   Domain          19..43
FT                   /label= TM1
FT   Domain          44..63
FT                   /note= "Extracellular loop"
FT   Modified-site   47..52
FT                   /note= "N-myristoylated"
FT   Modified-site   51..54
FT                   /note= "N-glycosylated"
FT   Domain          64..86
FT                   /label= TM2
FT   Modified-site   71..76
FT                   /note= "N-myristoylated"
FT   Modified-site   81..86
FT                   /note= "N-myristoylated"
FT   Domain          87..94
FT                   /note= "Intracellular loop"
FT   Domain          95..117
FT                   /label= TM3
FT   Domain          118..234
FT                   /note= "Extracellular loop"
FT   Modified-site   171..174
FT                   /note= "N-glycosylated"
FT   Modified-site   183..188
FT                   /note= "N-myristoylated"
FT   Domain          235..256
FT                   /label= TM4
FT   Modified-site   240..245
FT                   /note= "N-myristoylated"
FT   Modified-site   252..257
FT                   /note= "N-myristoylated"
FT   Domain          257..270
FT                   /label= C-terminal_domain
XX
PN   WO200216603-A2.
XX
PD   28-FEB-2002.
XX
PF   21-AUG-2001; 2001WO-US41811.
XX
PR   21-AUG-2000; 2000US-226612P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Leiby KR;
XX
DR   WPI; 2002-315466/35.
DR   N-PSDB; AAI72633.
XX
PT   New tetraspanin 23228 polypeptide useful in screening assays,
PT   predictive medicine and as a prophylactic or therapeutic agent, e.g.,
PT   for hematopoietic and immune diseases such as diabetes or multiple
PT   sclerosis -
XX
PS   Claim 1; Page 113-14; 123pp; English.
XX
CC   This sequence shows tetraspanin 23228. This protein is a cell surface
CC   protein having four transmembrane domains (TM1-TM4).  TM1, TM3 and TM4
CC   have a single polar amino acid located within them, which may interact
CC   with each other and contribute to domain stability.  The cytoplasmic N-
CC   and C-terminal domains and the intracellular loop between TM2 and TM3
CC   are less than 30 amino acids in length.  The second extracellular domain
CC   between TM3 and TM4 contains conserved Cys residues and may function
CC   to bind extracellular growth factors, such as HB-EGF, TGF-alpha and
CC   amphiregulin. The 23228 protein is useful for diagnosing and treating
CC   23228-mediated disorders, e.g., haematopoietic and/or immune disorders
CC   such as diabetes mellitus, arthritis, multiple sclerosis,
CC   encephalomyelitis, dermatitis, Crohn's disease and asthma; cell
CC   proliferation and differentiation disorders, e.g., cancers or
CC   metastasis; and/or viral infections. 23228 polypeptide can control
CC   cellular signalling activity, bind to an extracellular growth factor,
CC   for example, amphiregulin, regulate cell proliferation, bind to a cell
CC   surface protein, to recruit intracellular kinases, to regulate cell
CC   motility, bind to another tetraspanin such as CD81, to associate with a
CC   B-Cell antigen receptor and the ability to modulate the association
CC   with a virus.
XX
SQ   Sequence   270 AA;

  Query Match             83.1%;  Score 1066;  DB 23;  Length 270;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-104;
  Matches  193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 4
AAB18969
ID   AAB18969 standard; Protein; 329 AA.
XX
AC   AAB18969;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of a human transmembrane protein.
XX
KW   Human; transmembrane protein; cell proliferation disorder;  myeloma;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW   allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW   Alzheimer's disease; Tourette's disorder.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   51
FT                   /note= "potential glycosylation site"
FT   Modified-site   171
FT                   /note= "potential glycosylation site"
FT   Modified-site   269
```

```
FT                  /note= "potential glycosylation site"
FT   Modified-site  285
FT                  /note= "potential phosphorylation site"
FT   Modified-site  325
FT                  /note= "potential phosphorylation site"
XX
PN   WO200056891-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07817.
XX
PR   22-MAR-1999;   99US-0125537.
PR   16-JUN-1999;   99US-0139565.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Yue H,  Lal P,  Tang YT,  Hillman JL,  Reddy R,  Bandman O,  Baughn MR;
PI   Lu DAM,  Azimzai Y,  Yang J;
XX
DR   WPI; 2000-579485/54.
DR   N-PSDB; AAA96482.
XX
PT   New human transmembrane proteins are used to treat a disease or
PT   condition associated with decreased expression of functional HTMP e.g.
PT   Tourette's disorder, angina and leukaemia -
XX
PS   Claim 1; Page 91; 130pp; English.
XX
CC   The present sequence represents a human transmembrane proteins (HTMP).
CC   Agonists and antagonists of the protein are used to treat a disease
CC   or condition associated with overexpression of the protein. Diseases
CC   and conditions which can be treated include cell proliferative,
CC   immunological, reproductive, smooth muscle and neurological disorders
CC   e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC   syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC   stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC   polynucleotides may be used to detect and quantify gene expression in
CC   biopsied tissues where protein expression may be correlated with disease
CC   e.g. to determine absence, presence or excess expression of HTMP or to
CC   monitor regulation of HTMP expression during therapeutic intervention.
XX
SQ   Sequence   329 AA;

  Query Match             83.1%;  Score 1066;  DB 21;  Length 329;
  Best Local Similarity   99.5%;  Pred. No. 1.6e-104;
  Matches  193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194


RESULT 5
ABB90429
ID   ABB90429 standard; Protein; 329 AA.
XX
AC   ABB90429;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2805.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US16450.
XX
PR   19-MAY-2000; 2000US-205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   N-PSDB; ABL90838.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 11; SEQ ID NO 2805; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   329 AA;

  Query Match             82.2%;  Score 1055;  DB 23;  Length 329;
  Best Local Similarity   98.5%;  Pred. No. 2.3e-103;
  Matches  191;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNIXALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LXPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||||||||
Db        181 RCGVPFSCCVRDPA 194
```

```
RESULT 6
ABP69599
ID   ABP69599 standard; Protein; 330 AA.
XX
AC   ABP69599;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1646.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   N-PSDB; ABZ11816.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders -
XX
PS   Claim 9; SEQ ID NO 1646; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   330 AA;

  Query Match             75.8%;  Score 972.5;  DB 23;  Length 330;
  Best Local Similarity   91.9%;  Pred. No. 1.3e-94;
  Matches  181;  Conservative    1;  Mismatches   10;  Indels    5;  Gaps    2;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60

Qy         61 LDPVWLFVVVGGVMSVLGFAGC---IGALRENTFLLKFFSVFLGLIFFLELATGILAFVF 117
              ||||||  | |      :|   |   ||||||||||||||||||||||||||||||||||
Db         61 LDPVWL--VCGSWRRHVGAGLCWAAIGALRENTFLLKFFSVFLGLIFFLELATGILAFVF 118

Qy        118 KDWIRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNP 177
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 KDWIRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNP 178

Qy        178 SRERCGVPFSCCVRDPA 194
              |||||||||||||||||
Db        179 SRERCGVPFSCCVRDPA 195


RESULT 7
AAB53438
ID   AAB53438 standard; Protein; 271 AA.
XX
AC   AAB53438;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen protein sequence SEQ ID NO:978.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   N-PSDB; AAC98195.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 11; Page 1550-1551; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence   271 AA;
```

```
  Query Match             69.0%;  Score 885;  DB 21;  Length 271;
  Best Local Similarity   79.9%;  Pred. No. 2e-85;
  Matches  155;  Conservative   16;  Mismatches   21;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| ||||||||::|||||
Db          4 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 61

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
               ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         62 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 121

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::|||||||||||||| |||| |||| | :||||||||||| | |||
Db        122 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 181

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||| :|||
Db        182 RCGVPFSCCTKDPA 195


RESULT 8
AAB49505
ID   AAB49505 standard; Protein; 268 AA.
XX
AC   AAB49505;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Clone HE8EJ16 #1.
XX
KW   Human; 4 transmembrane superfamily receptor protein;
KW   endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW   reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW   bacterial infection; viral; fungal.
XX
OS   Homo sapiens.
XX
PN   WO200070076-A1.
XX
PD   23-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13504.
XX
PR   19-MAY-1999;   99US-0135122.
PR   03-JUN-1999;   99US-0137797.
PR   11-JUN-1999;   99US-0138573.
PR   18-AUG-1999;   99US-0149447.
PR   28-JAN-2000; 2000US-0178770.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Fan P,  Roschke V,  Shi Y,  Komatsoulis GA;
PI   Rosen CA;
XX
DR   WPI; 2001-007502/01.
DR   N-PSDB; AAC90015.
XX
PT   Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT   superfamily receptor protein, useful for diagnosing, treating and/or
PT   preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS   Claim 11; Page 282; 297pp; English.
XX
CC   The present invention relates to isolated nucleic acids and proteins
CC   encoding human soluble 4 transmembrane superfamily receptor protein (see
CC   AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
CC   protein. The present protein can be used to screen for binding partners
CC   and molecules which modify its activity. Antibodies specific for the
CC   present protein can be used to treat and/or prevent diseases associated
CC   with abberant expression or activity of the present protein e.g.
CC   endocrine disorders e.g. Addison's disease, (cardio)vascular diseases
CC   e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural
CC   disorders e.g. Alzheimer's and Parkinson's disease, reproductive
CC   disorders, skin disorders e.g. psoriasis, renal system disorders e.g.
CC   nephritis, (auto)immune system disorders e.g. graft vs. host disease,
CC   hyperproliferative disorders e.g. neoplasms of the pancreas, ocular
CC   disorders e.g. glaucoma and infections caused by bacteria, viruses and
CC   fungi.
XX
SQ   Sequence   268 AA;

  Query Match             61.7%;  Score 791;  DB 22;  Length 268;
  Best Local Similarity   72.2%;  Pred. No. 1.8e-75;
  Matches  140;  Conservative   16;  Mismatches   36;  Indels    2;  Gaps    1;

Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| ||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
               ||||||:|||||| :|||||||||     :      ||      |  ||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGAYGKTLSFSSFFLCSWELFSFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::|||||||||||||| |||| |||| | :||||||||||| | |||
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRE 178

Qy        181 RCGVPFSCCVRDPA 194
              ||||||||| :|||
Db        179 RCGVPFSCCTKDPA 192


RESULT 9
AAY42381
ID   AAY42381 standard; Protein; 193 AA.
XX
AC   AAY42381;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Amino acid sequence of dk329_1.
XX
KW   secreted protein; cDNA library; clone; transmembrane protein;
KW   signal sequence cloning; hybridization cloning; gene therapy;
KW   receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         71..83
FT                   /label= Leader/Signal peptide
FT   Protein         84..193
FT                   /label= dk329_1 mature protein
XX
PN   WO9942470-A1.
XX
PD   26-AUG-1999.
XX
PF   18-FEB-1999;   99WO-US03458.
XX
PR   18-FEB-1998;   98US-0075038.
PR   17-FEB-1999;   99US-0251600.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Merberg D;
PI   Treacy M,  Agostino MJ,  Steininger RJ;
XX
DR   WPI; 1999-518580/43.
DR   N-PSDB; AAZ20854.
XX
PT   New polynucleotides encoding human secreted proteins used for
PT   therapeutic, diagnostic and research purposes.
XX
```

PS Claim 13; Page 101-102; 125pp; English.
XX
CC This is the amino acid sequence of the dk329_1 protein, which is
CC derived from the dk329_1 clone isolated from a human fetal kidney cDNA
CC library.
CC The PNs and proteins of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC hematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumor invasion suppressor activity, and tumor
CC inhibition activity. The PNs are also stated to be useful for gene
CC therapy. Other activities include inhibiting the growth, infection or
CC function of bacteria, fungi, viruses and other parasites; effecting
CC bodily characteristics such as, e.g. weight, color, skin, etc.,
CC effecting biorhythms or caricadic cycles; enhancing fertility; treatment
CC of depression; treatment of pain; hormonal or endocrine activity.
XX
SQ Sequence 193 AA;

Query Match 54.5%; Score 699; DB 20; Length 193;
Best Local Similarity 62.9%; Pred. No. 7.1e-66;
Matches 139; Conservative 16; Mismatches 30; Indels 36; Gaps 4;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| |||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::||||||||||||||| |||                       | :
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEY----------------------IPMQV 155

Qy        181 RCGVPFSCCVRDPA-------MSSTPSVAMMSGSNWSWSSR 214
                | |      ||       |||| ||||| | |   :||
Db        156 ESDVAF----HSPAALKIPQKMSSTLSVAMMPGKNQKLTSR 192
```

RESULT 10
AAU39073
ID AAU39073 standard; Protein; 193 AA.
XX
AC AAU39073;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human secreted protein dk329_1.
XX
KW Human; secreted protein; antiinflammatory; immunosuppressive;
KW nootropic; neuroprotective; antiarthritic; antimicrobial; vulnerary;
KW cytostatic; antidiabetic; virucide; antiinfertility; anticonvulsant;
KW vasotropic; antiparkinsonian; immunostimulant; dermatological;
KW antirheumatic; antitumor; antiulcer; osteopathic; tranquiliser;
KW cerebroprotective; cytokine; cell proliferation; cell differentiation;
KW immune deficiency; severe combined immunodeficiency; SCID; tumour;
KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;
KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; infection; cardiac disease;
KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;
KW food supplement; vaccine.
XX
OS Homo sapiens.
XX
PN WO200175068-A2.
XX
PD 11-OCT-2001.
XX
PF 22-MAR-2001; 2001WO-US09369.
XX
PR 30-MAR-2000; 2000US-0539330.
PR 04-DEC-2000; 2000US-0729674.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;
PI Clark H, Fechtel K, Merberg D;
XX
DR WPI; 2001-639363/73.
DR N-PSDB; AAS59291.
XX
PT Secreted human proteins, useful as vaccine for treating various
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and
PT nervous system disorders (e.g. stroke) -
XX
PS Disclosure; Page 577; 619pp; English.
XX
CC The invention relates to novel human secreted proteins, the nucleic
CC acids encoding them. The protein may exhibit cytokine, cell proliferation
CC or cell differentiation activity or may induce production of other
CC cytokines in certain cell populations and may exhibit immune stimulating
CC or immune suppressing activity, which is useful for the treatment of
CC various immune deficiencies and disorders e.g. severe combined
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,
CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation. The proteins are also useful in the treatment of diseases
CC and disorders including tissue, skin and organ transplantation and in
CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,
CC in the treatment of burns, incisions and ulcers; as well as in treatment
CC of periodontal disease, osteoporosis or osteoarthritis, mediated by
CC inflammatory processes, diseases of the peripheral nervous system,
CC Alzheimer's, Parkinson's disease, Huntington's disease,
CC amylotrophic lateral sclerosis, and Shy-Drager syndrome, infections,
CC infarction of cardiac and central nervous system vessel e.g. stroke,
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The
CC protein, having activin- or inhibin-related activities is useful as a
CC contraceptive based on the ability of inhibins to decrease fertility in
CC female mammals and decrease spermatogenesis in male mammals. The
CC proteins and nucleic acids are also useful as food supplements. The
CC present sequence represents a secreted protein of the invention.
XX
SQ Sequence 193 AA;

Query Match 54.5%; Score 699; DB 22; Length 193;
Best Local Similarity 62.9%; Pred. No. 7.1e-66;
Matches 139; Conservative 16; Mismatches 30; Indels 36; Gaps 4;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || |||||| |||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::||||||||||||||| |||                       | :
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEY----------------------IPMQV 155

Qy        181 RCGVPFSCCVRDPA-------MSSTPSVAMMSGSNWSWSSR 214
                | |      ||       |||| ||||| | |   :||
Db        156 ESDVAF----HSPAALKIPQKMSSTLSVAMMPGKNQKLTSR 192
```

RESULT 11
ABB55782
ID ABB55782 standard; Protein; 193 AA.

XX
AC ABB55782;
XX
DT 14-FEB-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 170.
XX
KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;
KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerary;
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;
KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory.
XX
OS Homo sapiens.
XX
PN US2001039335-A1.
XX
PD 08-NOV-2001.
XX
PF 04-DEC-2000; 2000US-0729674.
XX
PR 10-APR-1997; 97US-126425P.
PR 04-DEC-1997; 97US-067454P.
PR 20-DEC-1997; 97US-068379P.
PR 02-JAN-1998; 98US-070346P.
PR 07-JAN-1998; 98US-070643P.
PR 08-JAN-1998; 98US-070755P.
PR 13-JAN-1998; 98US-071304P.
PR 22-JAN-1998; 98US-072134P.
PR 30-JAN-1998; 98US-073095P.
PR 18-FEB-1998; 98US-075038P.
PR 30-MAR-2000; 2000US-0539330.
PR 23-NOV-1998; 98US-0197886.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (AGOS/) AGOSTINO M J.
PA (STEI/) STEININGER R J.
PA (SPAU/) SPAULDING V.
PA (WONG/) WONG G G.
PA (CLAR/) CLARK H.
PA (FECH/) FECHTEL K.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark H, Fechtel K;
XX
DR WPI; 2002-040725/05.
DR N-PSDB; ABA90960.
XX
PT New secreted proteins and encoding polynucleotides, useful in gene
PT therapies, particularly for preventing or treating autoimmune
PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,
PT stroke or inflammations -
XX
PS Disclosure; Page 307-308; 349pp; English.
XX
CC The invention relates to isolated polynucleotides (ABA90876-ABA90968 and
CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially
CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and
CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in
CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1
CC are deposited with the American Type Culture Collection (ATCC) with
CC accession number 98599. The polynucleotides and encoded polypeptides have
CC cytostatic, anti-inflammatory, immunomodulator, vulnerary,
CC neuroprotective, activin, inhibin, chemotactic, hAemostatic, thrombolytic
CC and anti-inflammatory activity and acting as cytokine modulators,
CC haematopoiesis regulators, tissue growth modulators and/or cadherin
CC suppressors. The polypeptides and polynucleotides are useful in gene
CC therapies, particularly for preventing, treating or ameliorating any of
CC the following diseases: immune deficiency and disorders; e.g. bacterial
CC or fungal infections, autoimmune disorders, cancer, systemic lupus
CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell
CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or
CC osteoarthritis; central and peripheral nervous system diseases and
CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's
CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;
CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis
CC or systemic inflammatory response syndrome, ischaemia-reperfusion
CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or
CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus
CC foliaceus.
XX
SQ Sequence 193 AA;

Query Match 54.5%; Score 699; DB 23; Length 193;
Best Local Similarity 62.9%; Pred. No. 7.1e-66;
Matches 139; Conservative 16; Mismatches 30; Indels 36; Gaps 4;

```
Qy          1 MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGG 60
              | ||  |:: ||| || |||:||||::|| ||  || ||||| |||||||||::|||||
Db          1 MSGK--HYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEKGVLSNISSITDLGG 58

Qy         61 LDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDW 120
              ||||||:|||||| :|||||||||||||||||||||||||:||||||  |:||||||||
Db         59 FDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDW 118

Qy        121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRE 180
              |:||| ||||||::|||||||||||||| |||                       | :
Db        119 IKDQLYFFINNNIRAYRDDIDLQNLIDFTQEY----------------------IPMQV 155

Qy        181 RCGVPFSCCVRDPA-------MSSTPSVAMMSGSNWSWSSR 214
                | |      ||       |||| ||||| | |   :||
Db        156 ESDVAF----HSPAALKIPQKMSSTLSVAMMPGKNQKLTSR 192
```

RESULT 12
ABP98695
ID ABP98695 standard; Protein; 512 AA.
XX
AC ABP98695;
XX
DT 13-JUN-2003 (first entry)
XX
DE Rat gamma-hydroxybutyrate receptor.
XX
KW rat; gamma-hydroxybutyrate; anti-epileptic; anxiolytic;
KW antineurodegeneration; antipsychotic; brain; dopamine; opioid; GABA;
KW gamma-aminobutyric acid; diagnosis; epilepsy; anxiety; sleep disorder;
KW behavioral disorder; neurodegeneration; Parkinson's disease; psychosis;
KW schizophrenia.
XX
OS Rattus rattus.
XX
PN WO200078948-A2.
XX
PD 28-DEC-2000.
XX
PF 19-JUN-2000; 2000WO-FR01687.
XX
PR 18-JUN-1999; 99FR-0007784.
XX
PA (UYPA-) UNIV PASTEUR LOUIS.
XX
PI Andriamanpandry C, Maitre M;
XX
DR WPI; 2001-091570/10.

```
DR   N-PSDB; ACC44092.
XX
PT   New rat brain gamma-hydroxybutyrate receptor and its encoding nucleic
PT   acid, useful for identifying agents for treating e.g. epilepsy -
XX
PS   Claim 1; Page 62-63; 66pp; French.
XX
CC   This sequence represents a novel rat gamma-hydroxybutyrate receptor
CC   (GHBR) isolated from rat brain hypocampal cell cDNA library. The
CC   invention relates to the isolation of this novel sequence, fragments of
CC   it and homologues of the sequence except for those homologues that having
CC   GenBank accession numbers AAC 17120 (human tetraspan NET-4), AA615405,
CC   AA967250 or AI467230. GHBR is the receptor for gamma-hydroxybutyrate in
CC   the rat brain and is involved in regulating dopaminergic, opioid and
CC   GABA(gamma-aminobutyric acid)ergic activities. The nucleic acid that
CC   encodes the protein is used: (1) as primers or probes for detection and
CC   amplification, particularly for screening gene libraries to identify
CC   promoters and regulators of the GHBR gene; (2) for expression of
CC   recombinant polypeptides; and (3) to detect allelic variants, mutations,
CC   deletions, loss of heterozygosity or genetic abnormalities in the GHBR
CC   gene (for diagnosing diseases, or susceptibility, associated with
CC   abnormal expression of GHBR). The protein (or cells and transgenic
CC   animals expressing it) is used: (1) to screen for agents that interact
CC   with GHBR; (2) to study expression/activity of the receptor, including
CC   its interaction with other compounds; and (3) to raise antibodies (Ab)
CC   specific for GHBR. The Ab are used to detect/measure GHBR in diagnostic
CC   immunoassays. The agents are used to prevent or treat diseases associated
CC   with abnormal expression or activity of GHBR, particularly those
CC   involving cerebral GABA(gamma-aminobutyric acid)ergic and/or dopaminergic
CC   activities, e.g. epilepsy, anxiety, sleep or behavioral disorders,
CC   withdrawal from addictive drugs, neurodegeneration (Parkinson's disease),
CC   psychoses, schizophrenia and regulation of secretion of hormones (growth
CC   hormone and prolactin) that are under dopaminergic control.
XX
SQ   Sequence   512 AA;

  Query Match             49.9%;  Score 640.5;  DB 22;  Length 512;
  Best Local Similarity   89.6%;  Pred. No. 3.9e-59;
  Matches  120;  Conservative    3;  Mismatches   10;  Indels    1;  Gaps    1;

Qy         59 GGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFK 118
              |||||||||||:||:|||||||||||||||||||||||||||||||||||| ||||||||
Db         97 GGLDPVWLFVVIGGIMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELAAGILAFVFK 156

Qy        119 DWIRDQLNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFN-CTDLNP 177
              |||||||| ||||||||||||||||||||||||||||||||||||||  :     ||
Db        157 DWIRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIRTSTALTSNP 216

Qy        178 SRERCGVPFSCCVR 191
              ||||||||| | ||
Db        217 SRERCGVPFFCWVR 230


RESULT 13
AAY76266
ID   AAY76266 standard; Protein; 270 AA.
XX
AC   AAY76266;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human secreted protein encoded by gene 10 fragment.
XX
KW   Human; secreted protein; cancer; tumour; developmental abnormality;
KW   foetal deficiency; blood disorder; immune system disorder; inflammation;
KW   autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW   schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW   atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW   digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
KW   therapy; chromosome 10.
XX
OS   Homo sapiens.
XX
PN   WO9958660-A1.
XX
PD   18-NOV-1999.
XX
PF   06-MAY-1999;   99WO-US09847.
XX
PR   12-MAY-1998;   98US-0085093.
PR   12-MAY-1998;   98US-0085094.
PR   12-MAY-1998;   98US-0085105.
PR   12-MAY-1998;   98US-0085180.
PR   18-MAY-1998;   98US-0085906.
PR   18-MAY-1998;   98US-0085920.
PR   18-MAY-1998;   98US-0085921.
PR   18-MAY-1998;   98US-0085922.
PR   18-MAY-1998;   98US-0085923.
PR   18-MAY-1998;   98US-0085924.
PR   18-MAY-1998;   98US-0085928.
PR   18-MAY-1998;   98US-0085925.
PR   18-MAY-1998;   98US-0085927.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Florence K,  Ni J,  Rosen CA,  Carter KC,  Moore PA;
PI   Olsen HS,  Shi Y,  Young PE,  Wei F,  Brewer LA,  Soppet DR;
PI   Lafleur DW,  Endress GA,  Ebner R;
XX
DR   WPI; 2000-062296/05.
DR   N-PSDB; AAZ65350.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders  -
XX
PS   Disclosure; Page 428-429; 475pp; English.
XX
CC   AAZ65250 to AAZ65350 represent 97 isolated human secreted protein genes.
CC   AAY76124 to AAY76223 are the secreted proteins encoded by the 97 human
CC   genes. The gene encoding this protein was found to be on chromosome 10.
CC   The genes and their corresponding secreted polypeptides are
CC   useful for preventing, treating or ameliorating medical conditions,
CC   e.g. by protein or gene therapy. Also pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new genes. Specific
CC   uses are described for each of the 97 genes, based on which tissues they
CC   are most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, tumours, developmental abnormalities
CC   and foetal deficiencies, blood disorders, diseases of the immune system,
CC   autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC   disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC   disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC   disorders, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
CC   The sequences shown in AAY76224 to AAY76424 represent fragments of the
CC   secreted proteins.
XX
SQ   Sequence   270 AA;

  Query Match             49.8%;  Score 638.5;  DB 21;  Length 270;
  Best Local Similarity   49.4%;  Pred. No. 2.9e-59;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :|||| ||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:|||||| || ||   : |||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
                 | :|:|:|||||||||||| |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182
```

```
Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :   | ||              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 14
AAB87034
ID   AAB87034 standard; Protein; 270 AA.
XX
AC   AAB87034;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Human secreted protein TANGO 339, SEQ ID NO:3.
XX
KW   Secreted protein; transmembrane protein; TANGO; human; drug screening;
KW   activity modulator; expression modulator; cancer; immunological disorder;
KW   cytostatic; immunomodulatory; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200109162-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-US20935.
XX
PR   30-JUL-1999;   99US-0365164.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Fraser CC,  Sharp JD,  Kirst SJ,  Barnes TM,  Wrighton N,  Myers PS;
PI   Pan Y;
XX
DR   WPI; 2001-138647/14.
DR   N-PSDB; AAF90628, AAF90629.
XX
PT   Nucleic acids encoding secreted TANGO and MANGO polypeptides, useful
PT   for the prevention, diagnosis and treatment of, e.g. cancers and immune
PT   disorders -
XX
PS   Claim 9; Page 220-221; 332pp; English.
XX
CC   The invention relates to novel secreted/transmembrane proteins, and
CC   nucleic acids encoding them. The novel proteins are designated TANGO 339,
CC   TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO
CC   393, TANGO 402, MANGO 346 and MANGO 349 and are of human origin, and a
CC   murine TANGO 393 is also included within the scope of the invention. The
CC   invention also encompasses fragments and variants of the proteins of the
CC   invention, and nucleic acids encoding them. The invention additionally
CC   relates to host cells comprising a nucleic acid of the invention; methods
CC   for the production of a protein of the invention; an antibody specific
CC   for a protein of the invention; methods for detecting a protein or
CC   nucleic acid of the invention; and methods of identifying agents which
CC   bind to or modulate the activity of a protein of the invention. The novel
CC   secreted proteins, nucleic acids encoding them, and antibodies against
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate expression or activity of the secreted
CC   proteins. The secreted proteins of the invention may also be used to
CC   identify modulators of expression or activity, which may be useful in
CC   the treatment of disorders associated with the proteins of the
CC   invention e.g., cancers and immunological disorders. The present
CC   sequence represents a human TANGO protein of the invention.
XX
SQ   Sequence   270 AA;

  Query Match             49.8%;  Score 638.5;  DB 22;  Length 270;
  Best Local Similarity   49.4%;  Pred. No. 2.9e-59;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :|||| ||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:|||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
               |  :|:|:|||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW-------------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :   | ||              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


RESULT 15
AAE18535
ID   AAE18535 standard; Protein; 270 AA.
XX
AC   AAE18535;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human tetraspan protein, TSPAN-7.
XX
KW   Human; tetraspan protein; TSPAN-7; hyperproliferative disorder;
KW   neoplastic disease; prostate cancer; antisense therapy; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200206340-A2.
XX
PD   24-JAN-2002.
XX
PF   13-JUL-2001; 2001WO-US22134.
XX
PR   14-JUL-2000; 2000US-218280P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Reinhard CJ,  Garcia PD;
XX
DR   WPI; 2002-179783/23.
DR   N-PSDB; AAD29623.
XX
PT   New inhibitor of tetraspan protein useful for reducing the expression
PT   or activities of tetraspan protein in cells, and for treating a
PT   hyperproliferative disorder, especially cancer in a mammal  -
XX
PS   Claim 17; Fig 3; 88pp; English.
XX
CC   The invention relates to inhibitors of tetraspan protein, TSPAN-7
CC   which are used to reduce or decrease the expression of TSPAN-7
CC   in a mammalian cell and for treating hyperproliferative disorder,
CC   especially cancer in a mammal. The invention also provides TSPAN-7
CC   polypeptides and polynucleotides. TSPAN-7 inhibitors are useful for
CC   inhibiting proliferation of mammalian cells, including tumour cells,
CC   for decreasing the side effects of cancer therapy and for treating
CC   neoplastic diseases. They are used to modulate TSPAN-7 expression
CC   and function in cancer cells, particularly in prostate cancer cells.
CC   They are also used in antisense therapy. The present sequence is
CC   human TSPAN-7 protein.
XX
SQ   Sequence   270 AA;

  Query Match             49.8%;  Score 638.5;  DB 23;  Length 270;
```

```
  Best Local Similarity   49.4%;  Pred. No. 2.9e-59;
  Matches  119;  Conservative   37;  Mismatches   66;  Indels   19;  Gaps    2;

Qy          5 HQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWGEKGVLSNISALTDLGGLDPV 64
              :  :   :| |  || || :||:||: | :|| :||||| ||||||::: :| : |:|||
Db          3 YYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEKGVLSDLTKVTRMHGIDPV 62

Qy         65 WLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELATGILAFVFKDWIRDQ 124
               | ::|| ||  ||||||:||||||  || ||   : ||||||||  :|||:|:||:||:
Db         63 VLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIFFLELAVAVLAFLFQDWVRDR 122

Qy        125 LNFFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIYFNCTDLNPSRERCGV 184
                 |  :|:|:||||||||||||  |:   |||| || ||:||:||||:  : |||:|||
Db        123 FREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPEDWDLNVYFNCSGASYSREKCGV 182

Qy        185 PFSCCVRDPAMSSTPS-----VAMMSGSNW----------SWSSRAPYTPKAVWAS 225
              |||||| |||     :     | :  | |              ||  |  |    |: :
Db        183 PFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFTKGCIQALESWLPRNIYIVAGVFIA 242

Qy        226 L 226
              :
Db        243 I 243


Search completed: November 21, 2003, 13:43:36
Job time : 42 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 15:29:38 ; Search time 335 Seconds
(without alignments)
2273.109 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972970/runat_21112003_125336_27261/app_query.fasta_1.391
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09972970_@CGN_1_1_221_@runat_21112003_125336_27261
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1283	100.0	2538	10	US-09-972-970-2	Sequence 2, Appli
2	1215.5	94.7	813	10	US-09-934-268-3	Sequence 3, Appli
3	1215.5	94.7	813	14	US-10-162-435-37	Sequence 37, Appl
4	1215.5	94.7	2461	12	US-10-252-157-153	Sequence 153, App
5	1215.5	94.7	3184	10	US-09-934-268-1	Sequence 1, Appli
6	1215.5	94.7	3184	14	US-10-162-435-35	Sequence 35, Appl
7	1212.5	94.5	1126	14	US-10-103-196-4	Sequence 4, Appli
8	958.5	74.7	1655	9	US-09-925-299-205	Sequence 205, App
9	958.5	74.7	1655	11	US-09-925-299-205	Sequence 205, App
10	918.5	71.6	1174	14	US-10-103-196-5	Sequence 5, Appli
11	892.5	69.6	1178	14	US-10-103-196-10	Sequence 10, Appl
12	807	62.9	1110	9	US-09-729-674-169	Sequence 169, App
13	801.5	62.5	1988	11	US-09-822-846-126	Sequence 126, App
14	790	61.6	864	10	US-09-875-440-1	Sequence 1, Appli
15	654.5	51.0	1388	11	US-09-905-674-1	Sequence 1, Appli
16	654.5	51.0	2715	11	US-09-796-753-125	Sequence 125, App
17	641.5	50.0	2672	11	US-09-892-877-20	Sequence 20, Appl
18	641.5	50.0	2672	11	US-09-948-783-20	Sequence 20, Appl
19	635.5	49.5	1428	11	US-09-991-053-15	Sequence 15, Appl
20	635.5	49.5	1428	11	US-09-957-187-15	Sequence 15, Appl
21	583.5	45.5	1667	11	US-09-892-877-266	Sequence 266, App
22	583.5	45.5	1667	11	US-09-948-783-268	Sequence 268, App
23	537	41.9	368	13	US-10-042-417-37	Sequence 37, Appl
c 24	490.5	38.2	839	14	US-10-106-698-2019	Sequence 2019, Ap
25	476.5	37.1	1932	11	US-09-984-271-41	Sequence 41, Appl
26	427.5	33.3	816	11	US-09-991-053-78	Sequence 78, Appl
27	427.5	33.3	816	11	US-09-957-187-78	Sequence 78, Appl
28	334.5	26.1	1778	11	US-09-946-374-122	Sequence 122, App
29	334.5	26.1	1778	12	US-10-015-387A-122	Sequence 122, App
30	334.5	26.1	1778	12	US-10-006-130A-122	Sequence 122, App
31	334.5	26.1	1778	12	US-10-199-672-277	Sequence 277, App
32	334.5	26.1	1778	12	US-10-006-172A-122	Sequence 122, App
33	334.5	26.1	1778	12	US-10-187-749-277	Sequence 277, App
34	334.5	26.1	1778	12	US-10-194-457-277	Sequence 277, App
35	334.5	26.1	1778	12	US-10-184-642-277	Sequence 277, App
36	334.5	26.1	1778	12	US-10-196-747-277	Sequence 277, App
37	334.5	26.1	1778	12	US-10-015-392A-122	Sequence 122, App
38	334.5	26.1	1778	12	US-10-017-253A-122	Sequence 122, App
39	334.5	26.1	1778	12	US-10-173-689-277	Sequence 277, App
40	334.5	26.1	1778	12	US-10-173-690-277	Sequence 277, App
41	334.5	26.1	1778	12	US-10-173-691-277	Sequence 277, App
42	334.5	26.1	1778	12	US-10-173-692-277	Sequence 277, App
43	334.5	26.1	1778	12	US-10-173-694-277	Sequence 277, App
44	334.5	26.1	1778	12	US-10-173-698-277	Sequence 277, App
45	334.5	26.1	1778	12	US-10-173-699-277	Sequence 277, App

ALIGNMENTS

RESULT 1
US-09-972-970-2
; Sequence 2, Application US/09972970
; Patent No. US20020164693A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: TM4SF Receptor Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT056P1
; CURRENT APPLICATION NUMBER: US/09/972,970
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11130
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/195,336
; PRIOR FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-972-970-2

Alignment Scores:
Pred. No.: 2.59e-139 Length: 2538
Score: 1283.00 Matches: 233

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-972-970-4 (1-233) x US-09-972-970-2 (1-2538)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTC 178

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 CTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTC 238

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGC 298

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCC 358

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 GGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTT 418

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGG 478

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 ATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATT 538

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGG 598

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAG 658

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrProSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 CGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTCAACACCCAGT 718

Qy        201 ValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778

Qy        221 AlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||||||||||||||||||||||||||||||||
Db        779 GCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 817

RESULT 2
US-09-934-268-3
; Sequence 3, Application US/09934268
; Patent No. US20020172986A1
; GENERAL INFORMATION:
;  APPLICANT: Leiby, Kevin R.
;  TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY
;  TITLE OF INVENTION:  MEMBER AND USES THEREOF
;  FILE REFERENCE: 10448-079001
;  CURRENT APPLICATION NUMBER: US/09/934,268
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 813
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-934-268-3

Alignment Scores:
Pred. No.:              3.73e-132      Length:        813
Score:                  1215.50        Matches:       227
Percent Similarity:     96.60%         Conservative:  0
Best Local Similarity:  96.60%         Mismatches:    6
Query Match:            94.74%         Indels:        2
DB:                     10             Gaps:          1

US-09-972-970-4 (1-233) x US-09-934-268-3 (1-813)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 60

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 120

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 180

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 240

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 300

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 360

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        361 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 420

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 480

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 540

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        541 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 600

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||||      |||||
Db        601 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 660

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   |||||||||||||||||||||||||||||||
Db        661 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 703

RESULT 3
US-10-162-435-37
; Sequence 37, Application US/10162435
; Publication No. US20030096305A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Curtis , Rory A. J.
;  APPLICANT:  Kapeller-Libermann, Rosana
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Leiby , Kevin R.
;  TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND
```

```
;  TITLE OF INVENTION:  CELL SURFACE PROTEIN FAMILY MEMBERS
;  FILE REFERENCE: 10448-189001
;  CURRENT APPLICATION NUMBER: US/10/162,435
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US 09/836,499
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: PCT/US01/12420
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: US 60/197,507
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: US 09/891,008
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: PCT/US01/19963
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/214,220
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 09/860,868
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: PCT/US01/16013
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,674
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/886,429
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: PCT/US01/20055
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/213,963
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 10/041,406
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: PCT/US02/00275
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/260,286
;  PRIOR FILING DATE: 2001-01-08
;  PRIOR APPLICATION NUMBER: US 09/934,268
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/41811
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
;   LENGTH: 813
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-162-435-37

Alignment Scores:
Pred. No.:              3.73e-132       Length:         813
Score:                  1215.50         Matches:        227
Percent Similarity:     96.60%          Conservative:   0
Best Local Similarity:  96.60%          Mismatches:     6
Query Match:            94.74%          Indels:         2
DB:                     14              Gaps:           1

US-09-972-970-4 (1-233) x US-10-162-435-37 (1-813)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 60

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 120

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 180

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 240

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 300

Qy        101 GlyLeuIlePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 360

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        361 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 420

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 480

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 540

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        541 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 600

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||  ||||||||||||||||||||||||||||||||||  |||||
Db        601 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 660

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||  ||||||||||||||||||||||||||||||||
Db        661 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 703

RESULT 4
US-10-252-157-153
; Sequence 153, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
;  APPLICANT: Faris, Mary
;  APPLICANT:  Pearson, Cecelia I.
;  TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
;  FILE REFERENCE: PA-0027-1 US
;  CURRENT APPLICATION NUMBER: US/10/252,157
;  CURRENT FILING DATE:  2002-10-01
;  PRIOR APPLICATION NUMBER: 60/295,048
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 501
;  SOFTWARE: PERL Program
; SEQ ID NO 153
;   LENGTH: 2461
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030190640A1 234056.5
US-10-252-157-153

Alignment Scores:
Pred. No.:              1.89e-131       Length:         2461
Score:                  1215.50         Matches:        227
Percent Similarity:     96.60%          Conservative:   0
Best Local Similarity:  96.60%          Mismatches:     6
Query Match:            94.74%          Indels:         2
DB:                     12              Gaps:           1

US-09-972-970-4 (1-233) x US-10-252-157-153 (1-2461)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 173

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        174 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 233

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        234 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 293

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 353

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 413

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 473

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        474 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 533

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 593

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        594 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 653

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        654 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 713

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||      |||||
Db        714 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 773

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        774 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 816

RESULT 5
US-09-934-268-1
; Sequence 1, Application US/09934268
; Patent No. US20020172986A1
; GENERAL INFORMATION:
;  APPLICANT: Leiby, Kevin R.
;  TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY
;  TITLE OF INVENTION:  MEMBER AND USES THEREOF
;  FILE REFERENCE: 10448-079001
;  CURRENT APPLICATION NUMBER: US/09/934,268
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3184
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (168)...(977)
US-09-934-268-1

Alignment Scores:
Pred. No.:              2.75e-131      Length:          3184
Score:                  1215.50        Matches:         227
Percent Similarity:     96.60%         Conservative:    0
Best Local Similarity:  96.60%         Mismatches:      6
Query Match:            94.74%         Indels:          2
DB:                     10             Gaps:            1

US-09-972-970-4 (1-233) x US-09-934-268-1 (1-3184)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 227

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 287

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 347

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 407

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 467

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 527

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        528 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 587

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 647

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 707

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        708 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 767

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||      |||||
Db        768 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 827

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        828 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 870

RESULT 6
US-10-162-435-35
; Sequence 35, Application US/10162435
; Publication No. US20030096305A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Curtis , Rory A. J.
;  APPLICANT:  Kapeller-Libermann, Rosana
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Leiby , Kevin R.
;  TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND
;  TITLE OF INVENTION:  CELL SURFACE PROTEIN FAMILY MEMBERS
;  FILE REFERENCE: 10448-189001
;  CURRENT APPLICATION NUMBER: US/10/162,435
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US 09/836,499
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: PCT/US01/12420
;  PRIOR FILING DATE: 2001-04-17
;  PRIOR APPLICATION NUMBER: US 60/197,507
;  PRIOR FILING DATE: 2000-04-18
```

```
;  PRIOR APPLICATION NUMBER: US 09/891,008
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: PCT/US01/19963
;  PRIOR FILING DATE: 2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/214,220
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 09/860,868
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: PCT/US01/16013
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,674
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/886,429
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: PCT/US01/20055
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/213,963
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 10/041,406
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: PCT/US02/00275
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/260,286
;  PRIOR FILING DATE: 2001-01-08
;  PRIOR APPLICATION NUMBER: US 09/934,268
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/41811
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/226,612
;  PRIOR FILING DATE: 2000-08-21
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 3184
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (168)...(977)
US-10-162-435-35

Alignment Scores:
Pred. No.:                2.75e-131      Length:          3184
Score:                    1215.50        Matches:         227
Percent Similarity:       96.60%         Conservative:    0
Best Local Similarity:    96.60%         Mismatches:      6
Query Match:              94.74%         Indels:          2
DB:                       14             Gaps:            1

US-09-972-970-4 (1-233) x US-10-162-435-35 (1-3184)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 227

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 287

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 347

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 407

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 467

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 527

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        528 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 587

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 647

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 707

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        708 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 767

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||||      |||||
Db        768 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 827

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||
Db        828 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 870

RESULT 7
US-10-103-196-4
; Sequence 4, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE: PT008P1
;  CURRENT APPLICATION NUMBER: US/10/103,196
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/707,936
;  PRIOR FILING DATE: 2000-11-08
;  PRIOR APPLICATION NUMBER: PCT/US00/13504
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: 60/178,770
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/149,447
   PRIOR FILING DATE: 1999-08-18
;  PRIOR APPLICATION NUMBER: 60/138,573
;  PRIOR FILING DATE: 1999-06-11
;  PRIOR APPLICATION NUMBER: 60/137,797
;  PRIOR FILING DATE: 1999-06-03
;  PRIOR APPLICATION NUMBER: 60/135,122
;  PRIOR FILING DATE: 1999-05-19
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1126
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-103-196-4

Alignment Scores:
Pred. No.:                1.35e-131      Length:          1126
Score:                    1212.50        Matches:         226
Percent Similarity:       96.60%         Conservative:    1
Best Local Similarity:    96.17%         Mismatches:      6
Query Match:              94.51%         Indels:          2
DB:                       14             Gaps:            1

US-09-972-970-4 (1-233) x US-10-103-196-4 (1-1126)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 216

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
```

```
Db        217 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 276

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60

Db        277 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 336

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80

Db        337 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 396

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100

Db        397 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 456

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120

Db        457 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 516

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140

Db        517 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 576

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160

Db        577 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 636

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180

Db        637 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 696

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199

Db        697 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 756

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219

Db        757 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 816

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233

Db        817 AAAGGCTGCGTGGGCCAGTTTGAGAAATGGCTGCAGGACAACC 859

RESULT 8
US-09-925-299-205
; Sequence 205, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 205
;   LENGTH: 1655
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1548)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1559)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1564)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1623)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1643)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-205

Alignment Scores:
Pred. No.:              1.09e-101       Length:         1655
Score:                  958.50          Matches:        180
Percent Similarity:     84.19%          Conservative:   17
Best Local Similarity:  76.92%          Mismatches:     33
Query Match:            74.71%          Indels:         4
DB:                     9               Gaps:           2

US-09-972-970-4 (1-233) x US-09-925-299-205 (1-1655)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20

Db        277 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 330

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40

Db        331 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 390

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60

Db        391 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 450

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80

Db        451 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 510

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100

Db        511 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 570

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120

Db        571 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 630

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140

Db        631 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 690

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160

Db        691 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 750

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180

Db        751 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAG 810

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199

Db        811 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 870

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219

Db        871 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 930

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232

Db        931 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 970

RESULT 9
US-09-925-299-205
; Sequence 205, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
```

```
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 205
;   LENGTH: 1655
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1548)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1559)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1564)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1623)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (1643)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-205

Alignment Scores:
Pred. No.:                1.09e-101       Length:          1655
Score:                    958.50          Matches:         180
Percent Similarity:       84.19%          Conservative:    17
Best Local Similarity:    76.92%          Mismatches:      33
Query Match:              74.71%          Indels:          4
DB:                       11              Gaps:            2

US-09-972-970-4 (1-233) x US-09-925-299-205 (1-1655)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||   ||||||  |||||||||
Db        277 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 330

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||||::::::||||||   ||||||      ||||||  |||||||||
Db        331 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 390

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   ||||||||||||||||||||||||||||::::::|||||||||||||||
Db        391 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 450

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
                ||||||||||||||||||:::||||||||||||||||||   :::|||||||||||||
Db        451 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 510

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 570

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||      |||:::||||||||||||||||||||||||
Db        571 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 630

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::|||||||||   ||||||||||||||||||:::::||||||||||||||||||
Db        631 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 690

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||   ||||||||||||   ||||||||||||   |||
Db        691 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 750

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              :::||||||||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        751 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAG 810

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||   :::||||||||||    ||||||||||||
Db        811 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 870

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||||||||   |||   |||          :::||||||          |||
Db        871 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 930

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||      ||||||||||||||||||||||||
Db        931 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 970

RESULT 10
US-10-103-196-5
; Sequence 5, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
;  FILE REFERENCE: PT008P1
;  CURRENT APPLICATION NUMBER: US/10/103,196
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/707,936
;  PRIOR FILING DATE: 2000-11-08
;  PRIOR APPLICATION NUMBER: PCT/US00/13504
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: 60/178,770
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/149,447
;  PRIOR FILING DATE: 1999-08-18
;  PRIOR APPLICATION NUMBER: 60/138,573
;  PRIOR FILING DATE: 1999-06-11
;  PRIOR APPLICATION NUMBER: 60/137,797
;  PRIOR FILING DATE: 1999-06-03
;  PRIOR APPLICATION NUMBER: 60/135,122
;  PRIOR FILING DATE: 1999-05-19
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1174
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (2)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-103-196-5

Alignment Scores:
Pred. No.:                3.09e-97        Length:          1174
Score:                    918.50          Matches:         179
Percent Similarity:       83.40%          Conservative:    17
Best Local Similarity:    76.17%          Mismatches:      34
Query Match:              71.59%          Indels:          6
DB:                       14              Gaps:            2

US-09-972-970-4 (1-233) x US-10-103-196-5 (1-1174)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||   ||||||  |||||||||
Db         93 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 146

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||||::::::||||||   ||||||      ||||||  |||||||||
Db        147 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 206
```

```
Qy          41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db         207 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 266

Qy          61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
Db         267 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 326

Qy          81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
Db         327 GGGTGCATTGGAGCGTA-CGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 385

Qy         101 GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120
Db         386 GGAATTATTTTCTTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTG 445

Qy         120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIl 140
Db         446 GATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACAT 505

Qy         140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGl 160
Db         506 TGATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGG 565

Qy         160 yProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGl 180
Db         566 AGCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGA 625

Qy         180 uArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrP 199
Db         626 GCGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACAC 685

Qy         199 roSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrP 219
Db         686 TCAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACAC 745

Qy         219 roLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
Db         746 GAAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 786
```

RESULT 11
US-10-103-196-10
; Sequence 10, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT008P1
; CURRENT APPLICATION NUMBER: US/10/103,196
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/707,936
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/13504
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/178,770
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/149,447
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/138,573
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/137,797
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/135,122
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1178
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-196-10

Alignment Scores:

Pred. No.:	3.37e-94	Length:	1178
Score:	892.50	Matches:	179
Percent Similarity:	82.70%	Conservative:	17
Best Local Similarity:	75.53%	Mismatches:	34
Query Match:	69.56%	Indels:	8
DB:	14	Gaps:	2

US-09-972-970-4 (1-233) x US-10-103-196-10 (1-1178)

```
Qy           1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
Db          91 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 144

Qy          21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db         145 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 204

Qy          41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db         205 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 264

Qy          61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
Db         265 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 324

Qy          81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSer-ValPheLe 100
Db         325 GGGTGCATTGGAGCGTA-CGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTCGTGTTCCT 383

Qy         100 u-GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAsp 119
Db         384 CGGGAATTATTTTCTTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGAC 443

Qy         120 TrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAsp 139
Db         444 TGGATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGAC 503

Qy         140 IleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArg 159
Db         504 ATTGATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTT 563

Qy         160 GlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArg 179
Db         564 GGAGCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGA 623

Qy         180 GluArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerTh 198
Db         624 GAGCGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAAC 683

Qy         198 rProSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrTh 218
Db         684 ACTCAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTAC 743

Qy         218 rProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
Db         744 ACGAAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 786
```

RESULT 12
US-09-729-674-169
; Sequence 169, Application US/09729674
; Patent No. US20010039335A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Spaulding, Vikki

```
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6055-64X
;  CURRENT APPLICATION NUMBER: US/09/729,674
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: 09/539,330
;  PRIOR FILING DATE: 2000-03-30
;  NUMBER OF SEQ ID NOS: 283
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 169
;   LENGTH: 1110
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-729-674-169

Alignment Scores:
Pred. No.:              2.99e-84        Length:         1110
Score:                  807.00          Matches:        163
Percent Similarity:     76.50%          Conservative:   16
Best Local Similarity:  69.66%          Mismatches:     30
Query Match:            62.90%          Indels:         26
DB:                     9               Gaps:           3

US-09-972-970-4 (1-233) x US-09-729-674-169 (1-1110)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||    ||||||   |||||||||
Db        176 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 229

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||||:::::||||||   ||||||       ||||||   |||||||||
Db        230 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 289

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   ||||||||||||||||||||||||||:::::||||||||||||||||
Db        290 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 349

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||:::||||||||||||||||||   :::||||||||||||||
Db        350 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 409

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 469

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||
Db        470 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 529

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::|||||||||   ||||||||||||||||||:::::|||||||||||||||||||
Db        530 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 589

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||   |||||||||   |||
Db        590 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATAT-TCC----------------- 630

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                                                     |||   |||||||||
Db        631 ---------------------------------------------AATGCAAGTCGAGAG 645

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||||||||||   :::|||||||||    ||||||||||||
Db        646 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 705

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||||||   |||  |||        :::||||||          |||
Db        706 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 765

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
```

```
                  |||||||||      |||||||||||||||||||||||
Db        766 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 805

RESULT 13
US-09-822-846-126
; Sequence 126, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Steininger II, Robert J.
;  APPLICANT:  Bowman, Michael R.
;  APPLICANT:  Spaulding, Vikki
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6400
;  CURRENT APPLICATION NUMBER: US/09/822,846
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,605
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 629
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 126
;   LENGTH: 1988
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-846-126

Alignment Scores:
Pred. No.:              3.07e-83        Length:         1988
Score:                  801.50          Matches:        153
Percent Similarity:     85.94%          Conservative:   12
Best Local Similarity:  79.69%          Mismatches:     25
Query Match:            62.47%          Indels:         2
DB:                     11              Gaps:           1

US-09-972-970-4 (1-233) x US-09-822-846-126 (1-1988)

Qy         43 TrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAsp 62
              |||   |||||||||||||||||||||||||:::::|||||||||||||||   |||
Db          2 TGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGAC 61

Qy         63 ProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCys 82
              |||||||||||||:::|||||||||||||||||   :::|||||||||||||||||||
Db         62 CCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGC 121

Qy         83 IleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeu 102
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        122 ATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATT 181

Qy        103 IlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArg 122
              ||||||||||||||||||||      |||:::|||||||||||||||||||||||:::
Db        182 ATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAA 241

Qy        123 AspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAspLeu 142
              |||||||||   ||||||||||||||||||:::::|||||||||||||||||||||||
Db        242 GACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTG 301

Qy        143 GlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGlyProAsn 162
              ||||||||||||||||||   ||||||||||||   ||||||||||||   |||   :::
```

```
Db        302 CAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGAT 361

Qy        163 AspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGluArgCys 182

Db        362 GATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGT 421

Qy        183 GlyValProPheSerCysCysValArgAspProAla----MetSerSerThrProServa 201

Db        422 GGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGT 481

Qy        201 lAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAl 221

Db        482 GGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGC 541

Qy        221 aValTrpAlaSerLeuArgSerGlyCysArgThr 232

Db        542 TGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 575

RESULT 14
US-09-875-440-1
; Sequence 1, Application US/09875440
; Patent No. US20020156035A1
; GENERAL INFORMATION:
;  APPLICANT: Reinhard, Christoph
;  APPLICANT:  Jefferson, Anne B.
;  APPLICANT:  Winter, Jill A.
;  APPLICANT:  Randazzo, Filippo
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
;  TITLE OF INVENTION:  NEOPLASTIC DISEASE USING NET-4 MODULATORS
;  FILE REFERENCE: PP-01701.002/200130.522
;  CURRENT APPLICATION NUMBER: US/09/875,440
;  CURRENT FILING DATE:  2001-06-05
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 864
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: NET-4 oligonucleotide used in cell proliferation
;   OTHER INFORMATION: assay on SW620 cells
US-09-875-440-1

Alignment Scores:
Pred. No.:              2e-82       Length:         864
Score:                  790.00      Matches:        139
Percent Similarity:     88.51%      Conservative:   15
Best Local Similarity:  79.89%      Mismatches:     18
Query Match:            61.57%      Indels:         2
DB:                     10          Gaps:           1

US-09-972-970-4 (1-233) x US-09-875-440-1 (1-864)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20

Db        348 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 401

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40

Db        402 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 461

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60

Db        462 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 521

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80

Db        522 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 581

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100

Db        582 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 641

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120

Db        642 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 701

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140

Db        702 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 761

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160

Db        762 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 821

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174

Db        822 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGAT 863

RESULT 15
US-09-905-674-1
; Sequence 1, Application US/09905674
; Publication No. US20030039647A1
; GENERAL INFORMATION:
;  APPLICANT: Reinhard, Christoph
;  APPLICANT:  Garcia, Pablo
;  TITLE OF INVENTION: TETRASPAN PROTEIN AND USES THEREOF
;  FILE REFERENCE: PP-01700.002/200130.521
;  CURRENT APPLICATION NUMBER: US/09/905,674
;  CURRENT FILING DATE:  2001-07-13
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1388
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1285, 1377
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: 1285, 1377
;   OTHER INFORMATION: n = A,T,C or G
US-09-905-674-1

Alignment Scores:
Pred. No.:              2.67e-66    Length:         1388
Score:                  654.50      Matches:        125
Percent Similarity:     70.13%      Conservative:   37
Best Local Similarity:  54.11%      Mismatches:     67
Query Match:            51.01%      Indels:         2
DB:                     11          Gaps:           1

US-09-972-970-4 (1-233) x US-09-905-674-1 (1-1388)

Qy          5 HisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPheLeuPheGlyPhe 24

Db         73 TATTATAGATACTCTAACGCCAAAGTCAGCTGCTGGTACAAGTACCTCCTTTTCAGCTAC 132

Qy         25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44

Db        133 AACATCATCTTCTGGTTGGCTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGC 192

Qy         45 GluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProVal 64

Db        193 GAAAAGGGTGTGCTGTCCGACCTCACCAAAGTGACCCGGATGCATGGAATCGACCCTGTG 252

Qy         65 TrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84

Db        253 GTGCTGGTCCTGATGGTGGGCGTGGTGATGTTCACCCTGGGGTTCGCCGGCTGCGTGGGG 312

Qy         85 AlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeuIlePhe 104

Db        313 GCTCTGCGGGAGAATATCTGCTTGCTCAACTTTTTCTGTGGCACCATCGTGCTCATCTTC 372

Qy        105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArgAspGln 124
```

```
Db        373 TTCCTGGAGCTGGCTGTGGCCGTGCTGGCCTTCCTGTTCCAGGACTGGGTGAGGGACCGG 432

Qy        125 LeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAspLeuGlnAsn 144

Db        433 TTCCGGGAGTTCTTCGAGAGCAACATCAAGTCCTACCGGGACGATATCGATCTGCAAAAC 492

Qy        145 LeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGlyProAsnAspTrp 164

Db        493 CTCATCGACTCCCTTCAGAAAGCTAACCAGTGCTGTGGCGCATATGGCCCTGAAGACTGG 552

Qy        165 AsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGluArgCysGlyVal 184

Db        553 GACCTCAACGTCTACTTCAATTGCAGCGGTGCCAGCTACAGCCGAGAGAAGTGCGGGGTC 612

Qy        185 ProPheSerCysCysValArgAspProAlaMetSer---Ser-ThrProSerValAlaMe 203

Db        613 CCCTTCTCCTGCTGCGTGCCAGATCCTGCGCAAAAAGTTGTGAACACACAGTGTGGATAT 672

Qy        203 tMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaValTr 223

Db        673 GATGTCAGGATTCAGCTGAAGAGCAAGTGGGATGAGTCCATCTTCACGAAAGGCTGCATC 732

Qy        223 pAlaSerLeuArgSerGlyCysArgThrThr 233

Db        733 CAGGCGCTGGAAAGCTGGCTCCCGCGGAACA 763
```

Search completed: November 21, 2003, 17:20:02
Job time : 340 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 15:29:28 ; Search time 66 Seconds
(without alignments)
1558.216 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972970/runat_21112003_125336_27250/app_query.fasta_1.391
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972970_@CGN_1_1_56_@runat_21112003_125336_27250 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	476.5	37.1	1932	4	US-09-482-273-41	Sequence 41, Appl
2	266	20.7	399	3	US-09-188-930-62	Sequence 62, Appl
3	266	20.7	399	3	US-09-188-930-243	Sequence 243, App
4	266	20.7	399	4	US-09-312-283C-62	Sequence 62, Appl
5	266	20.7	399	4	US-09-312-283C-243	Sequence 243, App
6	250.5	19.5	827	3	US-09-333-599-5	Sequence 5, Appli
7	250.5	19.5	827	4	US-09-499-781-5	Sequence 5, Appli
8	250.5	19.5	870	3	US-09-333-599-1	Sequence 1, Appli
9	250.5	19.5	870	4	US-09-499-781-1	Sequence 1, Appli
10	249.5	19.4	1344	3	US-08-705-771-8	Sequence 8, Appli
11	248	19.3	1452	2	US-08-807-044-4	Sequence 4, Appli
12	248	19.3	1452	5	PCT-US91-04986-1	Sequence 1, Appli
13	248	19.3	1624	3	US-08-430-225A-19	Sequence 19, Appl
14	232	18.1	977	2	US-08-855-140-2	Sequence 2, Appli
15	232	18.1	977	4	US-09-016-434-938	Sequence 938, App
16	228	17.8	1151	2	US-08-807-044-2	Sequence 2, Appli
17	207.5	16.2	687	1	US-08-254-493-2	Sequence 2, Appli
18	207.5	16.2	687	1	US-08-408-222B-2	Sequence 2, Appli
19	207.5	16.2	1120	1	US-08-408-222B-3	Sequence 3, Appli
20	205.5	16.0	1120	1	US-08-254-493-3	Sequence 3, Appli
21	190	14.8	1001	3	US-08-705-771-6	Sequence 6, Appli
22	186	14.5	473	4	US-09-702-705-1565	Sequence 1565, Ap
23	186	14.5	473	4	US-09-736-457-1565	Sequence 1565, Ap
24	185.5	14.5	560	3	US-09-221-298-12	Sequence 12, Appl
25	183.5	14.3	933	3	US-08-808-148-2	Sequence 2, Appli
26	183.5	14.3	1289	3	US-09-020-956-111	Sequence 111, App
27	183.5	14.3	1289	3	US-09-030-607-111	Sequence 111, App
28	183.5	14.3	1289	4	US-09-439-313-111	Sequence 111, App
29	183.5	14.3	1289	4	US-09-352-616A-111	Sequence 111, App
30	183.5	14.3	1289	4	US-09-232-149A-111	Sequence 111, App
31	165	12.9	398	3	US-09-385-982-273	Sequence 273, App
c 32	155.5	12.1	801	3	US-09-020-956-16	Sequence 16, Appl
c 33	155.5	12.1	801	3	US-09-030-607-16	Sequence 16, Appl
c 34	155.5	12.1	801	4	US-09-439-313-16	Sequence 16, Appl
c 35	155.5	12.1	801	4	US-09-352-616A-16	Sequence 16, Appl
c 36	155.5	12.1	801	4	US-09-232-149A-16	Sequence 16, Appl
37	152	11.8	740	3	US-09-020-956-17	Sequence 17, Appl
38	152	11.8	740	3	US-09-030-607-17	Sequence 17, Appl
39	152	11.8	740	4	US-09-439-313-17	Sequence 17, Appl
40	152	11.8	740	4	US-09-352-616A-17	Sequence 17, Appl
41	152	11.8	740	4	US-09-232-149A-17	Sequence 17, Appl
42	134.5	10.5	1782	4	US-09-149-476-120	Sequence 120, App
43	118.5	9.2	1966	3	US-08-957-130-14	Sequence 14, Appl
c 44	117.5	9.2	751	3	US-09-020-956-12	Sequence 12, Appl
c 45	117.5	9.2	751	3	US-09-030-607-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-482-273-41
; Sequence 41, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: PZ030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1022)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-41

Alignment Scores:
Pred. No.: 2.75e-41 Length: 1932

```
Score:                  476.50        Matches:        104
Percent Similarity:     57.61%        Conservative:   36
Best Local Similarity:  42.80%        Mismatches:     70
Query Match:            37.14%        Indels:         33
DB:                     4             Gaps:           4

US-09-972-970-4 (1-233) x US-09-482-273-41 (1-1932)

Qy          3 GlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPheLeuPhe 22
              |||:::         |||     |||   |||             ||||||   ||||||
Db        133 GGGGAGGAGTTCTCCTTCGTCAGCCCGCTGGTG-----------AAATACCTGCTCTTC 180

Qy         23 GlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAla 42
                ||||||:::::|||||||||:::      :::   :::|||:::|||::::::|||
Db        181 TTCTTCAACATGCTCTTCTGGGTGATTTCCATGGTGATGGTGGCTGTGGGTGTCTACGCT 240

Qy         43 TrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAsp 62
                  |||          :::   ::::::   |||                 :::|||
Db        241 CGGCTAATGAAGCATGCAGAAGCAGCCCTAGCCTGCCTG-----------GCAGTGGAC 288

Qy         63 ProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCys 82
              |||      |||   :::|||||||||  :::|||   :::|||   |||   ||||||
Db        289 CCTGCCATCCTGCTGATCGTGGTGGGTGTCCTCATGTTCCTGCTCACCTTCTGTGGCTGC 348

Qy         83 IleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeu 102
              ||||||:::|||||||||||||      ||||||:::   ||||||:::   |||
Db        349 ATTGGGTCCCTCCGCGAGAACATCTGCCTCCTGCAGACGTTCTCCCTCTGCCTCACCGCT 408

Qy        103 IlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArg 122
              :::|||   |||:::||||||   |||||||||   |||||||||   |||       |||
Db        409 GTGTTCCTGCTGCAGCTGGCCGCTGGGATCCTGGGCTTCGTCTTCTCAGACAAGGCTCGA 468

Qy        123 AspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAspLeu 142
                 :::::::::      |||||||||   :::   ||||||||||:::|||||||
Db        469 GGGAAAGTGAGTGAGATCATCAACAATGCCATTGTGCACTACCGAGATGACTTGGATCTG 528

Qy        143 GlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGlyProAsn 162
              ||||||||||||||||||   |||:::   :::||||||||||||
Db        529 CAGAACCTCATTGATTTTGGCCAGAAAAAGTTTAGCTGCTGTGGAGGGATTTCCTACAAG 588

Qy        163 AspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGluArgCys 182
              ||||||:::   |||:::||||||||||||:::::   ||||||||||||||||||||||
Db        589 GACTGGTCTCAGAACATGTATTTCAACTGCTCAGAAGACAACCCCAGTCGAGAGCGCTGC 648

Qy        183 GlyValProPheSerCysCysValArgAspPro-------------------------- 193
                 ||||||:::||||||||:::      |||
Db        649 TCTGTGCCTTACTCCTGTTGCTTGCCTACTCCTGACCAGGCAGTGATCAACACTATGTGT 708

Qy        194 ------------------------------------AlaMetSerSerThrProServa 201
                                                   |||   ||||||||||||
Db        709 GGCCAAGGTATGCAGGCCTTTGACTACTTGGAAGCTAGCAAAGTCATCTACACCAATGGC 768

Qy        201 lAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAl 221
               :::   |||          ||||||           |||   |||:::
Db        769 TGTATTGACAAGT---------TGGTCAACTGGATACACAGCAACCTATTCTTACTTGGT 819

Qy        221 aValTrp 223
               ||||||
Db        820 GGTGTGG 826

RESULT 2
US-09-188-930-62
; Sequence 62, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 62
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: mouse
US-09-188-930-62

Alignment Scores:
Pred. No.:              8.51e-20      Length:         399
Score:                  266.00        Matches:        55
Percent Similarity:     54.76%        Conservative:   14
Best Local Similarity:  43.65%        Mismatches:     21
Query Match:            20.73%        Indels:         36
DB:                     3             Gaps:           4

US-09-972-970-4 (1-233) x US-09-188-930-62 (1-399)

Qy         11 ProGluValGlyCysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30
              |||||||||   ||||||   |||||||||:::|||||||||||||:::::||||||
Db          7 CCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTTGGTTT 66

Qy         31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSer 50
              ||||||      ||||||   ||||||||||||||||||   ||||||||||||||||||
Db         67 TTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCCTCTCC 126

Qy         51 AsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValVal 70
              |||||||||::::::||||||||||||||||   ||||||||||||||||||:::   |||
Db        127 AACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCTGAGTG 186

Qy         71 ---------------GlyGlyValMetSerValLeuGly---------------------- 78
                              |||       :::   |||
Db        187 GCCAGCCCGAGCCTGAGCTCTGTCAATGACATCCAAGGAGAAAATGAGGTTAATGAGAGA 246

Qy         79 ------------------------------------------------------PheAla 80
                                                                     ::::::
Db        247 CATTAATTAAACACTCCCTCACCCCACCGCACCAAACCAGTTGGGTTCTTCTGATATTCT 306

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePhe------SerVal 98
              |||   :::   |||:::      |||   :::   ||||||      :::
Db        307 GGAATACTCTGGGCTATG-----------TTTTATGTTTATTTCTTTTTTAATCGGTTG 354

Qy         99 PheLeuGlyLeuIlePhe 104
              :::   ||||||   |||
Db        355 TATTTTGGTCTTTTTTT 372

RESULT 3
US-09-188-930-243
; Sequence 243, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 243
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-188-930-243

Alignment Scores:
Pred. No.:              8.51e-20      Length:         399
```

```
Score:                  266.00          Matches:        55
Percent Similarity:     54.76%          Conservative:   14
Best Local Similarity:  43.65%          Mismatches:     21
Query Match:            20.73%          Indels:         36
DB:                     3               Gaps:           4

US-09-972-970-4 (1-233) x US-09-188-930-243 (1-399)

Qy         11 ProGluValGlyCysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30
              |||||||||   ||||||   |||||||||:::|||||||||||||:::::::||||||
Db          7 CCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTTGGTTT 66

Qy         31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSer 50
              ||||||      ||||||   |||||||||||||||||   ||||||||||||||||||
Db         67 TTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCCTCTCC 126

Qy         51 AsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValVal 70
              |||||||||::::::|||||||||||||||   |||||||||||||||||:::   |||
Db        127 AACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCTGAGTG 186

Qy         71 --------------GlyGlyValMetSerValLeuGly--------------------- 78
                          |||      :::   |||
Db        187 GCCAGCCCGAGCCTGAGCTCTGTCAATGACATCCAAGGAGAAAATGAGGTTAATGAGAGA 246

Qy         79 ----------------------------------------------------PheAla 80
                                                                  ::::::
Db        247 CATTAATTAAACACTCCCTCACCCCACCGCACCAAACCAGTTGGGTTCTTCTGATATTCT 306

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePhe------SerVal 98
              |||   :::   |||:::            |||   :::   ||||||        :::
Db        307 GGAATACTCTGGGCTATG-----------TTTTATGTTTATTTCTTTTTTAATCGGTTG 354

Qy         99 PheLeuGlyLeuIlePhe 104
              :::   ||||||   |||
Db        355 TATTTTGGTCTTTTTTT 372

RESULT 4
US-09-312-283C-62
; Sequence 62, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 62
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-312-283C-62

Alignment Scores:
Pred. No.:              8.51e-20        Length:         399
Score:                  266.00          Matches:        55
Percent Similarity:     54.76%          Conservative:   14
Best Local Similarity:  43.65%          Mismatches:     21
Query Match:            20.73%          Indels:         36
DB:                     4               Gaps:           4

US-09-972-970-4 (1-233) x US-09-312-283C-62 (1-399)

Qy         11 ProGluValGlyCysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30
              |||||||||   ||||||   |||||||||:::|||||||||||||:::::::||||||
Db          7 CCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTTGGTTT 66

Qy         31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSer 50
              ||||||      ||||||   |||||||||||||||||   ||||||||||||||||||
Db         67 TTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCCTCTCC 126

Qy         51 AsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValVal 70
              |||||||||::::::|||||||||||||||   |||||||||||||||||:::   |||
Db        127 AACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCTGAGTG 186

Qy         71 --------------GlyGlyValMetSerValLeuGly--------------------- 78
                          |||      :::   |||
Db        187 GCCAGCCCGAGCCTGAGCTCTGTCAATGACATCCAAGGAGAAAATGAGGTTAATGAGAGA 246

Qy         79 ----------------------------------------------------PheAla 80
                                                                  ::::::
Db        247 CATTAATTAAACACTCCCTCACCCCACCGCACCAAACCAGTTGGGTTCTTCTGATATTCT 306

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePhe------SerVal 98
              |||   :::   |||:::            |||   :::   ||||||        :::
Db        307 GGAATACTCTGGGCTATG-----------TTTTATGTTTATTTCTTTTTTAATCGGTTG 354

Qy         99 PheLeuGlyLeuIlePhe 104
              :::   ||||||   |||
Db        355 TATTTTGGTCTTTTTTT 372

RESULT 5
US-09-312-283C-243
; Sequence 243, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 243
;   LENGTH: 399
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-312-283C-243

Alignment Scores:
Pred. No.:              8.51e-20        Length:         399
Score:                  266.00          Matches:        55
Percent Similarity:     54.76%          Conservative:   14
Best Local Similarity:  43.65%          Mismatches:     21
Query Match:            20.73%          Indels:         36
DB:                     4               Gaps:           4

US-09-972-970-4 (1-233) x US-09-312-283C-243 (1-399)

Qy         11 ProGluValGlyCysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30
              |||||||||   ||||||   |||||||||:::|||||||||||||:::::::||||||
Db          7 CCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATGTCATATTTTGGTTT 66

Qy         31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSer 50
              ||||||      ||||||   |||||||||||||||||   ||||||||||||||||||
Db         67 TTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAAAAGGTGTCCTCTCC 126

Qy         51 AsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValVal 70
              |||||||||::::::|||||||||||||||   |||||||||||||||||:::   |||
Db        127 AACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCCTCTGAGTG 186

Qy         71 --------------GlyGlyValMetSerValLeuGly--------------------- 78
                          |||      :::   |||
```

```
Db        187 GCCAGCCCGAGCCTGAGCTCTGTCAATGACATCCAAGGAGAAAATGAGGTTAATGAGAGA 246

Qy         79 ------------------------------------------------------PheAla 80
                                                                     ::::::
Db        247 CATTAATTAAACACTCCCTCACCCCACCGCACCAAACCAGTTGGGTTCTTCTGATATTCT 306

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePhe------SerVal 98
               |||   :::   |||:::             |||   :::   ||||||        :::
Db        307 GGAATACTCTGGGCTATG-----------TTTTATGTTTATTTCTTTTTTAATCGGTTG 354

Qy         99 PheLeuGlyLeuIlePhe 104
              :::   ||||||   |||
Db        355 TATTTTGGTCTTTTTTT 372

RESULT 6
US-09-333-599-5
; Sequence 5, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
;  APPLICANT: Testa, Jacqueline E.
;  APPLICANT: Quigley, James P.
;  APPLICANT: Seandel, Marco
;  TITLE OF INVENTION: MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE: SUNY
;  CURRENT APPLICATION NUMBER: US/09/333,599
;  CURRENT FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 827
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (57)..(815)
US-09-333-599-5

Alignment Scores:
Pred. No.:              1.12e-17        Length:         827
Score:                  250.50          Matches:        77
Percent Similarity:     44.18%          Conservative:   33
Best Local Similarity:  30.92%          Mismatches:     99
Query Match:            19.52%          Indels:         40
DB:                     3               Gaps:           9

US-09-972-970-4 (1-233) x US-09-333-599-5 (1-827)

Qy          8 PheGlnGluProGluValGlyCys---------CysGlyLysTyrPheLeuPheGlyPhe 24
              |||   |||   :::      |||          |||   ||||||   ||||||   :::
Db         66 TTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTACCTGCTGTTTACCTAC 125

Qy         25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44
              |||      ||||||:::   |||          :::|||:::|||:::|||
Db        126 AATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGGCATCTGGACGCTGGCC 185

Qy         45 GluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProVal 64
                 |||                |||||   |||              |||
Db        186 CTCAAGAGT------GACTACATCAGCCTGCTGGCCTCAGGCACCTACCTGGCCACAGCC 239

Qy         65 TrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84
              ::::::   ||||||   |||   |||:::   |||   |||       ||||||
Db        240 TACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGTCTTGGGCTGCTGCGCC 299

Qy         85 AlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeuIlePhe 104
                 :::|||          ||||||:::   :::   :::   |||   :::||||||
Db        300 ACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCTGCTCCTCATCATCTTT 359

Qy        105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArgAspGln 124
                 ||||||:::      |||||||||||||:::   :::                 |||
Db        360 CTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTAC-----------TACCAGCAG 407

Qy        125 LeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAsp--------- 141
```

```
                 ||||||      :::      |||:::|||          |||   :::
Db        408 CTGAACACGGAGCTCAAGGAGAACCTGAAG---------GACACCATGACCAAGCGCTAC 458

Qy        142 --------------------LeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSer 154
                                  :::   :::   :::|||      |||:::   :::
Db        459 CACCAGCCGGGCCATGAGGCTGTGACCAGCGCTGTGGACCAGCTGCAGCAGGAGTTCCAC 518

Qy        155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174
              ||||||||||:::             ||||||      :::   :::
Db        519 TGCTGTGGCAGCAACAACTCACAGGACTGGCGAGACAGTGAGTGGATCCGC--------- 569

Qy        175 LeuAsnProSerArgGluArgCysGly------ValProPheSerCysCys--------- 189
                        |||:::|||      |||      ||||||   |||||||||
Db        570 ---------TCACAGGAGGCCGGTGGCCGTGTGGTCCCAGACAGCTGCTGCAAGACGGTG 620

Qy        190 ---------------ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204
                            ||||||   |||    :::||||||          |||
Db        621 GTGGCTCTTTGTGGACAGCGAGACCATGCCTCCAACATCTACAAGGTGGAGGGCGGCTGC 680

Qy        204 tSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaValTrpAl 224
              |||   |||   |||     |||||||||:::                  :::|||
Db        681 ATCACCAAGT---TGGAGACCTTCATCCAGGAGCACCTGAGGGTCATTGGGGCTGTGGGG 737

Qy        224 aSerLeuArgSerGlyCysArgThr 232
              |||             ||||||:::
Db        738 ATCGGCATTGCCTGTGTGCAGGTCT 762

RESULT 7
US-09-499-781-5
; Sequence 5, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
;  APPLICANT:  Testa, Jacqueline E.
;  APPLICANT:  Quigley, James P.
;  APPLICANT:  Seandel, Marco
;  TITLE OF INVENTION:  MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION:  ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE:  SUNY
;  CURRENT APPLICATION NUMBER: US/09/499,781
;  CURRENT FILING DATE:  2000-02-08
;  PRIOR APPLICATION NUMBER:  09/333,599
;  PRIOR FILING DATE:  1999-06-15
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 827
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (57)..(815)
US-09-499-781-5

Alignment Scores:
Pred. No.:              1.12e-17        Length:         827
Score:                  250.50          Matches:        77
Percent Similarity:     44.18%          Conservative:   33
Best Local Similarity:  30.92%          Mismatches:     99
Query Match:            19.52%          Indels:         40
DB:                     4               Gaps:           9

US-09-972-970-4 (1-233) x US-09-499-781-5 (1-827)

Qy          8 PheGlnGluProGluValGlyCys---------CysGlyLysTyrPheLeuPheGlyPhe 24
              |||   |||   :::      |||          |||   ||||||   ||||||   :::
Db         66 TTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTACCTGCTGTTTACCTAC 125

Qy         25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44
              |||      ||||||:::   |||          :::|||:::|||:::|||
Db        126 AATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGGCATCTGGACGCTGGCC 185

Qy         45 GluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProVal 64
                 |||                |||||   |||              |||
```

```
Db        186 CTCAAGAGT------GACTACATCAGCCTGCTGGCCTCAGGCACCTACCTGGCCACAGCC 239

Qy         65 TrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84

Db        240 TACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGTCTTGGGCTGCTGCGCC 299

Qy         85 AlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeuIlePhe 104

Db        300 ACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCTGCTCCTCATCATCTTT 359

Qy        105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArgAspGln 124

Db        360 CTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTAC-----------TACCAGCAG 407

Qy        125 LeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAsp--------- 141

Db        408 CTGAACACGGAGCTCAAGGAGAACCTGAAG---------GACACCATGACCAAGCGCTAC 458

Qy        142 --------------------LeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSer 154

Db        459 CACCAGCCGGGCCATGAGGCTGTGACCAGCGCTGTGGACCAGCTGCAGCAGGAGTTCCAC 518

Qy        155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174

Db        519 TGCTGTGGCAGCAACAACTCACAGGACTGGCGAGACAGTGAGTGGATCCGC--------- 569

Qy        175 LeuAsnProSerArgGluArgCysGly------ValProPheSerCysCys--------- 189

Db        570 ---------TCACAGGAGGCCGGTGGCCGTGTGGTCCCAGACAGCTGCTGCAAGACGGTG 620

Qy        190 ---------------ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204

Db        621 GTGGCTCTTTGTGGACAGCGAGACCATGCCTCCAACATCTACAAGGTGGAGGGCGGCTGC 680

Qy        204 tSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaValTrpAl 224

Db        681 ATCACCAAGT---TGGAGACCTTCATCCAGGAGCACCTGAGGGTCATTGGGGCTGTGGGG 737

Qy        224 aSerLeuArgSerGlyCysArgThr 232

Db        738 ATCGGCATTGCCTGTGTGCAGGTCT 762

RESULT 8
US-09-333-599-1
; Sequence 1, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
;  APPLICANT: Testa, Jacqueline E.
;  APPLICANT: Quigley, James P.
;  APPLICANT: Seandel, Marco
;  TITLE OF INVENTION: MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE: SUNY
;  CURRENT APPLICATION NUMBER: US/09/333,599
;  CURRENT FILING DATE: 1999-06-15
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-333-599-1

Alignment Scores:
Pred. No.:                1.21e-17      Length:          870
Score:                    250.50        Matches:         77
Percent Similarity:       44.18%        Conservative:    33
Best Local Similarity:    30.92%        Mismatches:      99
Query Match:              19.52%        Indels:          40
DB:                       3             Gaps:            9

US-09-972-970-4 (1-233) x US-09-333-599-1 (1-870)
```

```
Qy          8 PheGlnGluProGluValGlyCys---------CysGlyLysTyrPheLeuPheGlyPhe 24

Db        110 TTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTACCTGCTGTTTACCTAC 169

Qy         25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44

Db        170 AATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGGCATCTGGACGCTGGCC 229

Qy         45 GluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProVal 64

Db        230 CTCAAGAGT------GACTACATCAGCCTGCTGGCCTCAGGCACCTACCTGGCCACAGCC 283

Qy         65 TrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84

Db        284 TACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGTCTTGGGCTGCTGCGCC 343

Qy         85 AlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeuIlePhe 104

Db        344 ACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCTGCTCCTCATCATCTTT 403

Qy        105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArgAspGln 124

Db        404 CTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTAC-----------TACCAGCAG 451

Qy        125 LeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAsp--------- 141

Db        452 CTGAACACGGAGCTCAAGGAGAACCTGAAG---------GACACCATGACCAAGCGCTAC 502

Qy        142 --------------------LeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSer 154

Db        503 CACCAGCCGGGCCATGAGGCTGTGACCAGCGCTGTGGACCAGCTGCAGCAGGAGTTCCAC 562

Qy        155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174

Db        563 TGCTGTGGCAGCAACAACTCACAGGACTGGCGAGACAGTGAGTGGATCCGC--------- 613

Qy        175 LeuAsnProSerArgGluArgCysGly------ValProPheSerCysCys--------- 189

Db        614 ---------TCACAGGAGGCCGGTGGCCGTGTGGTCCCAGACAGCTGCTGCAAGACGGTG 664

Qy        190 ---------------ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204

Db        665 GTGGCTCTTTGTGGACAGCGAGACCATGCCTCCAACATCTACAAGGTGGAGGGCGGCTGC 724

Qy        204 tSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaValTrpAl 224

Db        725 ATCACCAAGT---TGGAGACCTTCATCCAGGAGCACCTGAGGGTCATTGGGGCTGTGGGG 781

Qy        224 aSerLeuArgSerGlyCysArgThr 232

Db        782 ATCGGCATTGCCTGTGTGCAGGTCT 806

RESULT 9
US-09-499-781-1
; Sequence 1, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
;  APPLICANT:  Testa, Jacqueline E.
;  APPLICANT:  Quigley, James P.
;  APPLICANT:  Seandel, Marco
;  TITLE OF INVENTION:  MONOCLONAL ANTIBOIDES THAT RECOGNIZE ANTIGENS
;  TITLE OF INVENTION:  ASSOCIATED WITH TUMOR METASTASIS
;  FILE REFERENCE:  SUNY
;  CURRENT APPLICATION NUMBER: US/09/499,781
;  CURRENT FILING DATE:  2000-02-08
;  PRIOR APPLICATION NUMBER:  09/333,599
;  PRIOR FILING DATE:  1999-06-15
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
US-09-499-781-1

Alignment Scores:
Pred. No.:                1.21e-17        Length:          870
Score:                    250.50          Matches:         77
Percent Similarity:       44.18%          Conservative:    33
Best Local Similarity:    30.92%          Mismatches:      99
Query Match:              19.52%          Indels:          40
DB:                       4               Gaps:            9

US-09-972-970-4 (1-233) x US-09-499-781-1 (1-870)

Qy          8 PheGlnGluProGluValGlyCys---------CysGlyLysTyrPheLeuPheGlyPhe 24
              |||   |||   :::      |||          |||  ||||||  ||||||   :::
Db        110 TTCAACGAGAAGAAGACAACATGTGGCACCGTTTGCCTCAAGTACCTGCTGTTTACCTAC 169

Qy         25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44
              |||      ||||||:::   |||        :::|||:::|||:::|||
Db        170 AATTGCTGCTTCTGGCTGGCTGGCCTGGCTGTCATGGCAGTGGGCATCTGGACGCTGGCC 229

Qy         45 GluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProVal 64
                 |||                 ||||||   |||                 |||
Db        230 CTCAAGAGT------GACTACATCAGCCTGCTGGCCTCAGGCACCTACCTGGCCACAGCC 283

Qy         65 TrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84
              ::::::   ||||||   |||   |||:::   |||   |||       ||||||
Db        284 TACATCCTGGTGGTGGCGGGCACTGTCGTCATGGTGACTGGGGTCTTGGGCTGCTGCGCC 343

Qy         85 AlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeuIlePhe 104
                 :::|||          ||||||:::   :::   :::   |||   :::||||||
Db        344 ACCTTCAAGGAGCGTCGGAACCTGCTGCGCCTGTACTTCATCCTGCTCCTCATCATCTTT 403

Qy        105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArgAspGln 124
              ||||||:::      ||||||||||||:::   :::                   |||
Db        404 CTGCTGGAGATCATCGCTGGTATCCTCGCCTACGCCTAC-----------TACCAGCAG 451

Qy        125 LeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAsp--------- 141
              ||||||      :::      |||:::|||          |||   :::
Db        452 CTGAACACGGAGCTCAAGGAGAACCTGAAG---------GACACCATGACCAAGCGCTAC 502

Qy        142 --------------------LeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSer 154
                                  :::   :::   :::|||      |||:::   :::
Db        503 CACCAGCCGGGCCATGAGGCTGTGACCAGCGCTGTGGACCAGCTGCAGCAGGAGTTCCAC 562

Qy        155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174
              |||||||||:::          ||||||      :::   :::
Db        563 TGCTGTGGCAGCAACAACTCACAGGACTGGCGAGACAGTGAGTGGATCCGC--------- 613

Qy        175 LeuAsnProSerArgGluArgCysGly------ValProPheSerCysCys--------- 189
                    |||:::|||      |||      ||||||   |||||||||
Db        614 ---------TCACAGGAGGCCGGTGGCCGTGTGGTCCCAGACAGCTGCTGCAAGACGGTG 664

Qy        190 ---------------ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204
                          ||||||   |||    :::||||||          |||
Db        665 GTGGCTCTTTGTGGACAGCGAGACCATGCCTCCAACATCTACAAGGTGGAGGGCGGCTGC 724

Qy        204 tSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaValTrpAl 224
              |||   |||   |||      |||||||||:::                 :::|||
Db        725 ATCACCAAGT---TGGAGACCTTCATCCAGGAGCACCTGAGGGTCATTGGGGCTGTGGGG 781

Qy        224 aSerLeuArgSerGlyCysArgThr 232
              |||             ||||||:::
Db        782 ATCGGCATTGCCTGTGTGCAGGTCT 806

RESULT 10
US-08-705-771-8
; Sequence 8, Application US/08705771
; Patent No. 6054289
;  GENERAL INFORMATION:
;    APPLICANT:  Paul Moore, Reiner Gentz, Hongjin Ji,
;    APPLICANT:  Jian Ni and Jing-Shan Hu
;    TITLE OF INVENTION:  Human Genes, Sequences and
;    TITLE OF INVENTION:  Expression Products
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/705,771
;      FILING DATE:  August 30, 1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MULLINS, J.G.
;      REGISTRATION NUMBER:  33,073
;      REFERENCE/DOCKET NUMBER:  325800-346 (PF196)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  973-994-1700
;      TELEFAX:  973-994-1744
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1344 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-705-771-8

Alignment Scores:
Pred. No.:                2.93e-17        Length:          1344
Score:                    249.50          Matches:         66
Percent Similarity:       44.98%          Conservative:    37
Best Local Similarity:    28.82%          Mismatches:      84
Query Match:              19.45%          Indels:          42
DB:                       3               Gaps:            9

US-09-972-970-4 (1-233) x US-08-705-771-8 (1-1344)

Qy         11 ProGluValGlyCysCysGly--------------------------LysTyrPheLeu 21
              ||||||:::      ||||||                          ||||||   :::
Db        123 CCAGAACTGAAGCGCTGCGGCATGGCGCGCGCCTGCCTCCAGGCCGTCAAGTACCTCATG 182

Qy         22 PheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrp 41
              |||   ||||||::::::||||||:::   |||          |||   :::|||:::|||
Db        183 TTCGCCTTCAACCTGCTCTTCTGGCTGGGAGGCTGTGGCGTGCTGGGTGTCGGCATCTGG 242

Qy         42 AlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeu 61
                    ::::::|||      :::   :::|||:::
Db        243 CTGGCCGCACAACAGGGGAGCTTTGCCACGCTGTCCTCT--------------TCCTTC 287

Qy         62 AspProValTrp---------LeuPheValValValGlyGlyValMetSerValLeuGly 78
                    :::|||          |||   ::::::   |||       :::       :::||||
Db        288 CCGTCCCTGTGGGCTGCCAACCTGCTCATCATCACCGGCGCCTTTGTCATGGCCATCGGC 347

Qy         79 PheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerVal 98
              |||   ||||||:::||||||::::::||||||       ||||||      |||   :::
Db        348 TTCGTGGGCTGCCTGGGTGCCATCAAGGAGAACAAGTGCCTCCTGCTCACTTTCTTCCTG 407

Qy         99 PheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLys 118
                 |||   |||:::|||   ||||||             ||||||   |||   :::
Db        408 CTGCTGCTGCTGGTGTTCCTGCTGGAGGGCACCATCGCCATCCTCTTCTTCGCCTACACG 467

Qy        119 AspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyr------ 136
                 |||   |||                  :::         :::       |||
Db        468 GACAAGATTGACAGGTATGCCCAGCAAGACCTGAAGAAAGGCTTGCACCTGTACGGCACG 527

Qy        137 ArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCys 156
```

```
Db        528 CAGGGCAACGTGGGCCTCACCAACGCCTGGAGCATCATCCAGACCGACTTCCGCTGCTGT 587

Qy        157 GlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsn 176

Db        588 GGCGTCTCCAACTACACTGACTGG---TTCGAGGTGTAC----------------AAC 626

Qy        177 ProSerArgGluArgCysGlyValProPheSerCysCys---------- 189

Db        627 GCCACGCGG-----------GTACCTGACTCCTGCTGCTTGGAGTTCAGTGAGAGCTGT 674

Qy        190 ---ValArgAspProAlaMet---------SerSerThrProSerValAlaMetMetSer 205

Db        675 GGGCTGCACGCCCCGGCACTGGTGGAGGGCCGTGCTACGAGAGGTGAAGGTGTGGCTTCA 734

Qy        206 Gly-SerAsnTrpSerTrpSerSer 213

Db        735 GGAGAACTGCTGGCTGTGGGCATCT 759

RESULT 11
US-08-807-044-4
; Sequence 4, Application US/08807044
; Patent No. 5863735
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/807,044
;      FILING DATE:  Herewith
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0224 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1452 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  180140
US-08-807-044-4

Alignment Scores:
Pred. No.:              4.75e-17      Length:          1452
Score:                  248.00        Matches:         67
Percent Similarity:     48.39%        Conservative:    38
Best Local Similarity:  30.88%        Mismatches:      81
Query Match:            19.33%        Indels:          31
DB:                     2             Gaps:            9
```

```
US-09-972-970-4 (1-233) x US-08-807-044-4 (1-1452)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGly------Lys 18

Db         44 GTGCCTGGAAAAGGGCAAGAATATCACGGCATGGGCATGAGTAGCTTGAAACTGCTGAAG 103

Qy         19 TyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIle 38

Db        104 TATGTCCTGTTTTTCTTCAACTTGCTCTTTTGGATCTGTGGCTGCTGCATTTTGGGCTTT 163

Qy         39 GlyLeuTrp------AlaTrpGlyGluLysGlyValLeu---SerAsnIleSerAlaLeu 55

Db        164 GGGATCTACCTGCTGATCCACAACAACTTCGGAGTGCTCTTCCATAACCTCCCCTCCCTC 223

Qy         56 ThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValValGlyGlyValMetSer 75

Db        224 ACG---CTGGGCAAT--------------GTGTTTGTCATCGTGGGCTCTATTATCATG 265

Qy         76 ValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhe 95

Db        266 GTAGTTGCCTTCCTGGGCTGCATGGGCTCTATCAAGGAAAACAAGTGTCTGCTTATGTCG 325

Qy         96 PheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPhe 115

Db        326 TTCTTCATCCTGCTGCTGATTATCCTCCTTGCTGAGGTGACCTTGGCCATCCTGCTCTTT 385

Qy        116 ValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAla 135

Db        386 GTATATGAACAGAAGCTGAATGAGTATGTGGCTAAGGGTCTGACCGACAGCATCCACCGT 445

Qy        136 TyrArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCys 155

Db        446 TACCACTCAGACAATAGCACCAAGGCAGCGTGGGACTCCATCCAGTCATTTCTGCAGTGT 505

Qy        156 CysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeu 175

Db        506 TGTGGTATAAATGGCACGAGTGATTGGACCAGTGGCCCACCAGCATCTTGC--------- 556

Qy        176 AsnProSer--------------------------ArgGluArg-CysGly---ValPr 185

Db        557 ---CCCTCAGATCGAAAAGTGGAGGGTTGCTATGCGAAAGCAAGACTGTGGTTTCATTCC 613

Qy        185 oPheSerCysCysValArgAspProAlaMetSerSerThrProSerVal 201

Db        614 AATTTCCTGT--------------ATATCGGAATCATCACCATCTGTG 647

RESULT 12
PCT-US91-04986-1
; Sequence 1, Application PC/TUS9104986
;  GENERAL INFORMATION:
;    APPLICANT:  Seed, Brian
;    APPLICANT:  Allen, Janet
;    APPLICANT:  Aruffo, Alejandro
;    APPLICANT:  Camerini, David
;    APPLICANT:  Lauffer, Leander
;    APPLICANT:  Oquendo, Carmen
;    APPLICANT:  Simmons, David L.
;    APPLICANT:  Stamenkovic, Ivan
;    APPLICANT:  Stengelin, Siegfried
;    APPLICANT:  Amiot, Martine
;    TITLE OF INVENTION:  Rapid Immunoselection Cloning Method
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee & Associates
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US91/04986
;       FILING DATE:  19910715
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/553,759
;       FILING DATE:  13-JUL-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/498,809
;       FILING DATE:  23-MAR-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/379,076
;       FILING DATE:  13-JUL-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/160,416
;       FILING DATE:  25-FEB-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Wall, Margaret M.
;       REGISTRATION NUMBER:  33,462
;       REFERENCE/DOCKET NUMBER:  11-88C
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  303-499-8080
;       TELEFAX:  303-499-8089
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1452 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  74..733
PCT-US91-04986-1

Alignment Scores:
Pred. No.:                4.75e-17        Length:          1452
Score:                    248.00          Matches:         67
Percent Similarity:       48.39%          Conservative:    38
Best Local Similarity:    30.88%          Mismatches:      81
Query Match:              19.33%          Indels:          31
DB:                       5               Gaps:            9

US-09-972-970-4 (1-233) x PCT-US91-04986-1 (1-1452)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGly------Lys 18
              :::|||||||||   |||   :::           :::                  |||
Db         44 GTGCCTGGAAAAGGGCAAGAATATCACGGCATGGGCATGAGTAGCTTGAAACTGCTGAAG 103

Qy         19 TyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIle 38
              |||   ||||||   ||||||::::::||||||:::   |||         |||
Db        104 TATGTCCTGTTTTTCTTCAACTTGCTCTTTTGGATCTGTGGCTGCTGCATTTTGGGCTTT 163

Qy         39 GlyLeuTrp------AlaTrpGlyGluLysGlyValLeu---SerAsnIleSerAlaLeu 55
              |||::::::                  |||||||||      |||:::   :::|||
Db        164 GGGATCTACCTGCTGATCCACAACAACTTCGGAGTGCTCTTCCATAACCTCCCCTCCCTC 223

Qy         56 ThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValValGlyGlyValMetSer 75
              |||   ||||||                 :::||||||:::||||||   ::::::
Db        224 ACG---CTGGGCAAT--------------GTGTTTGTCATCGTGGGCTCTATTATCATG 265

Qy         76 ValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhe 95
              |||:::   |||   ||||||:::|||:::::::::||||||      ||||||
Db        266 GTAGTTGCCTTCCTGGGCTGCATGGGCTCTATCAAGGAAAACAAGTGTCTGCTTATGTCG 325

Qy         96 PheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPhe 115
              |||   :::   |||   :::|||          |||:::          ||||||   |||
Db        326 TTCTTCATCCTGCTGCTGATTATCCTCCTTGCTGAGGTGACCTTGGCCATCCTGCTCTTT 385

Qy        116 ValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAla 135
              |||::::::      :::   :::   :::         :::   :::::::::
```

```
Db        386 GTATATGAACAGAAGCTGAATGAGTATGTGGCTAAGGGTCTGACCGACAGCATCCACCGT 445

Qy        136 TyrArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCys 155
              |||       |||          :::          |||       |||   :::      |||
Db        446 TACCACTCAGACAATAGCACCAAGGCAGCGTGGGACTCCATCCAGTCATTTCTGCAGTGT 505

Qy        156 CysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeu 175
              ||||||       |||   :::||||||                        :::|||
Db        506 TGTGGTATAAATGGCACGAGTGATTGGACCAGTGGCCCACCAGCATCTTGC--------- 556

Qy        176 AsnProSer--------------------------ArgGluArg-CysGly---ValPr 185
              ||||||                           :::   ||| ||||||   :::||
Db        557 ---CCCTCAGATCGAAAAGTGGAGGGTTGCTATGCGAAAGCAAGACTGTGGTTTCATTCC 613

Qy        185 oPheSerCysCysValArgAspProAlaMetSerSerThrProSerVal 201
              |   ||||||                        |||:::|||||||||
Db        614 AATTTCCTGT--------------ATATCGGAATCATCACCATCTGTG 647

RESULT 13
US-08-430-225A-19
; Sequence 19, Application US/08430225A
; Patent No. 6204000
;  GENERAL INFORMATION:
;    APPLICANT:  Dong, Jin-Tang; Barrett,
;    APPLICANT:  J. Carl; Lamb, Patricia W.; Isaacs, John T.
;    TITLE OF INVENTION:  DIAGNOSTIC METHODS AND
;    TITLE OF INVENTION:  GENE THERAPY USING REAGENTS DERIVED FROM THE
;    TITLE OF INVENTION:  HUMAN METASTASIS SUPPRESSOR GENE KAI1
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,225A
;      FILING DATE:  28-APR-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RICHARD W. BORK
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4172
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1624 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-430-225A-19

Alignment Scores:
Pred. No.:                5.6e-17         Length:          1624
Score:                    248.00          Matches:         74
Percent Similarity:       46.67%          Conservative:    38
Best Local Similarity:    30.83%          Mismatches:      83
Query Match:              19.33%          Indels:          45
DB:                       3               Gaps:            12

US-09-972-970-4 (1-233) x US-08-430-225A-19 (1-1624)

Qy         18 LysTyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAla 37
              ||||||||||||||   ||||||::::::|||::::::||||||||:::   |||
```

```
Db        193 AAATACTTTCTCTTCCTCTTCAACTTGATCTTCTTTATCCTGGGCGCAGTGATCCTGGGC 252

Qy         38 IleGlyLeuTrpAlaTrpGlyGluLys-----------GlyValLeuSerAsnIleSer 53
Db        253 TTCGGGGTGTGGATCCTGGCCGACAAGAGCAGTTTCATCTCTGTCCTGCAAACCTCCTCC 312

Qy         54 AlaLeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValValGlyGlyVal 73
Db        313 AGCTCGCTTAGGATGGGG-----------GCCTATGTCTTCATCGGCGTGGGGGCAGTC 360

Qy         74 MetSerValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeu 93
Db        361 ACTATGCTCATGGGCTTCCTGGGCTGCATCGGCGCCGTCAACGAGGTCCGCTGCCTGCTG 420

Qy         94 LysPhePheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeu 113
Db        421 GGGCTGTACTTTGCTTTCCTGCTCCTGATCCTCATTGCCCAGGTGACGGCCGGGGCCCTC 480

Qy        114 AlaPheValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnVal 133
Db        481 TTCTACTTCAACATGGGCAAGCTGAAGCAGGAGATGGGCGGCATCGTGACTGAGCTCATT 540

Qy        134 LysAlaTyr---------ArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGln 150
Db        541 CGAGACTACAACAGCAGTCGCGAGGAC---AGCCTGCAGGATGCCTGGGACTACGTGCAG 597

Qy        151 GluTyrTrpSerCysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPhe 170
Db        598 GCTCAGGTGAAGTGCTGCGGC-----------------TGG------GTCAGCTTCTAC 633

Qy        171 AsnCysThrAsp-----------LeuAsnProSerArgGluArgCysGlyValProPhe 186
Db        634 AACTGGACAGACAACGCTGAGCTCATGAAT------CGCCCTGAGGTCACCTACCCCTGT 687

Qy        187 SerCysCysValArg---------AspProAlaMetSer-----------------Ser 197
Db        688 TCCTGCGAAGTCAAGGGGGAAGAGGACAACAGCCTTTCTGTGAGGAAGGGCTTCTGCGAG 747

Qy        198 ThrProSerValAlaMetMetSerGlySer---------AsnTrpSerTrpSerSerArg 214
Db        748 GCCCCCGGCAACAGGACCCAGAGTGGCAACCACCCTGAGGACTGG--------------- 792

Qy        215 AlaProTyr-ThrProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
Db        793 ---CCTGTGTACCAGGAGGGCTGCATGGAGAAGGTGCAGGCGTGGCTGCAGGAGAACC 847
```

RESULT 14
US-08-855-140-2
; Sequence 2, Application US/08855140
; Patent No. 5854022
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,140
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0296 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MYOMNOT01
; CLONE: 779308
US-08-855-140-2

Alignment Scores:

Pred. No.:	1.36e-15	Length:	977
Score:	232.00	Matches:	60
Percent Similarity:	41.28%	Conservative:	30
Best Local Similarity:	27.52%	Mismatches:	92
Query Match:	18.08%	Indels:	36
DB:	2	Gaps:	6

US-09-972-970-4 (1-233) x US-08-855-140-2 (1-977)

```
Qy         15 CysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeu 34
Db        140 TGCTGCTTGAAGTACATGATGTTCCTCTTCAATTTGATATTCTGGCTCTGTGGCTGTGGG 199

Qy         35 PheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAla 54
Db        200 CTGCTGGGAGTGGGCATCTGGCTCTCCGTGTCCCAAGGCAACTTTGCCACCTTCTCC--- 256

Qy         55 LeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValValGlyGlyValMet 74
Db        257 ---CCCAGCTTCCCTTCGTTGTCTGCAGCCAACCTGGTCATCGCCATAGGCACCATTGTC 313

Qy         75 SerValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLys 94
Db        314 ATGGTGACGGGCTTCCTCGGCTGCCTGGGGGCCATCAAGGAAAACAAGTGCCTCCTCCTC 373

Qy         95 PhePheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAla 114
Db        374 AGCTTTTTCATCGTCCTGTTGGTCACCCTCCTAGCAGAGCTGATCTTACCCATCCTCTTC 433

Qy        115 PheValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLys 134
Db        434 TTTGTCTACATGGACAAGGTGAACGAGAACGCCAAGAAGGACCTGAAGGAAGGCCTGCTG 493

Qy        135 AlaTyr------ArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyr 152
Db        494 CTGTACCACACCGAGAACAACGTGGGGCTGAAGAACGCCTGGAACATCATCCAGGCTGAG 553

Qy        153 TrpSerCysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCys 172
Db        554 ATGCGATGCTGTGGTGTCACTGACTACACAGACTGG---------------------- 589

Qy        173 ThrAspLeuAsnProSerArgGluArgCysGlyValProPheSerCysCysValArgAsp 192
Db        590 ---------TACCCAGTGCTGGGGGAGAACACGGTTCCCGACCGCTGCTGCATG------ 634

Qy        193 ProAlaMetSerSerThrProSerValAlaMetMetSerGlySerAsnTrpSerTrpSer 212
Db        635 ----------------------------------------GAGAACTCCCAGGGCTGC 652

Qy        213 SerArgAlaProTyrThrProLysAlaValTrpAlaSerLeuArgSerGlyCys 230
```

```
                 |||         ||||||      :::|||         |||:::||||||
Db        653 GGGCGCAACGGCACCACGCCT------TTGTGG---------AGAACGGGCTGC 691
```

RESULT 15
US-09-016-434-938
; Sequence 938, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 938:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MYOMNOT01
; CLONE: 779308
US-09-016-434-938

Alignment Scores:

Pred. No.:	1.36e-15	Length:	977
Score:	232.00	Matches:	60
Percent Similarity:	41.28%	Conservative:	30
Best Local Similarity:	27.52%	Mismatches:	92
Query Match:	18.08%	Indels:	36
DB:	4	Gaps:	6

US-09-972-970-4 (1-233) x US-09-016-434-938 (1-977)

```
Qy         15 CysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeu 34
              ||||||  ||||||  :::|||  ||||||::::::||||||:::  |||
Db        140 TGCTGCTTGAAGTACATGATGTTCCTCTTCAATTTGATATTCTGGCTCTGTGGCTGTGGG 199

Qy         35 PheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAla 54
                 |||  :::|||:::|||          :::|||     :::      |||
Db        200 CTGCTGGGAGTGGGCATCTGGCTCTCCGTGTCCCAAGGCAACTTTGCCACCTTCTCC--- 256

Qy         55 LeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValValGlyGlyValMet 74
                          |||          |||  :::  :::|||  ::::::
Db        257 ---CCCAGCTTCCCTTCGTTGTCTGCAGCCAACCTGGTCATCGCCATAGGCACCATTGTC 313
```

```
Qy         75 SerValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLys 94
              |||  ||||||  ||||||:::||||||::::::||||||      ||||||
Db        314 ATGGTGACGGGCTTCCTCGGCTGCCTGGGGGCCATCAAGGAAAACAAGTGCCTCCTCCTC 373

Qy         95 PhePheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAla 114
              |||  :::  |||  :::          ||||||        ||||||
Db        374 AGCTTTTTCATCGTCCTGTTGGTCACCCTCCTAGCAGAGCTGATCTTACCCATCCTCTTC 433

Qy        115 PheValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLys 134
              ||||||:::  |||  :::  :::                :::        :::
Db        434 TTTGTCTACATGGACAAGGTGAACGAGAACGCCAAGAAGGACCTGAAGGAAGGCCTGCTG 493

Qy        135 AlaTyr------ArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyr 152
              |||        :::::::::  |||:::|||      :::      |||
Db        494 CTGTACCACACCGAGAACAACGTGGGGCTGAAGAACGCCTGGAACATCATCCAGGCTGAG 553

Qy        153 TrpSerCysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCys 172
                 |||||||||          ||||||
Db        554 ATGCGATGCTGTGGTGTCACTGACTACACAGACTGG---------------------- 589

Qy        173 ThrAspLeuAsnProSerArgGluArgCysGlyValProPheSerCysCysValArgAsp 192
                       |||                ||||||      ||||||:::
Db        590 ---------TACCCAGTGCTGGGGGAGAACACGGTTCCCGACCGCTGCTGCATG------ 634

Qy        193 ProAlaMetSerSerThrProSerValAlaMetMetSerGlySerAsnTrpSerTrpSer 212
                                                          |||
Db        635 ----------------------------------------GAGAACTCCCAGGGCTGC 652

Qy        213 SerArgAlaProTyrThrProLysAlaValTrpAlaSerLeuArgSerGlyCys 230
              |||        ||||||      :::|||         |||:::||||||
Db        653 GGGCGCAACGGCACCACGCCT------TTGTGG---------AGAACGGGCTGC 691
```

Search completed: November 21, 2003, 17:14:15
Job time : 70 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 15:29:18 ; Search time 2152 Seconds
(without alignments)
2631.479 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972970/runat_21112003_125335_27239/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972970_@CGN_1_1_2810_@runat_21112003_125335_27239 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1211.5	94.4	1201	9	AL528502	AL528502 AL528502
2	1169.5	91.2	1494	11	AK008761	AK008761 Mus muscu
3	1163.5	90.7	1194	9	AL529630	AL529630 AL529630
4	1158.5	90.3	1005	14	BY708665	BY708665 BY708665
5	1151.5	89.8	849	10	BE615772	BE615772 601279927
6	1151	89.7	938	13	BU856801	BU856801 AGENCOURT
7	1090.5	85.0	901	14	CA454987	CA454987 AGENCOURT
8	1087.5	84.8	781	12	BG770931	BG770931 602719148
9	1083	84.4	1032	12	BM563474	BM563474 AGENCOURT
10	1079.5	84.1	957	9	AL543914	AL543914 AL543914
11	1078.5	84.1	1201	9	AL556793	AL556793 AL556793
12	1071.5	83.5	1028	10	BG477727	BG477727 602521422
13	1066	83.1	718	12	BI914325	BI914325 603182305
14	1024.5	79.9	714	10	BG017161	BG017161 de61f11.y
15	1023	79.7	902	10	BG478644	BG478644 602525532
16	1020	79.5	970	12	BI909709	BI909709 603070671
17	1018.5	79.4	832	10	BF981395	BF981395 602309182
18	1017.5	79.3	782	12	BI158921	BI158921 602921828
19	1013.5	79.0	1020	12	BI557863	BI557863 603236858
20	1007	78.5	888	13	BX447619	BX447619 BX447619
21	967	75.4	654	12	BI829529	BI829529 603080360
22	966.5	75.3	588	10	BE615323	BE615323 601280719
23	963	75.1	594	10	BF043938	BF043938 BP250022B
24	954	74.4	883	14	CA986012	CA986012 AGENCOURT
25	944.5	73.6	3141	11	AK013350	AK013350 Mus muscu
26	935.5	72.9	753	14	CA315923	CA315923 UI-M-FW0-
27	930.5	72.5	899	13	BU911975	BU911975 AGENCOURT
28	905	70.5	538	10	BE683087	BE683087 181563 MA
29	905	70.5	842	13	BQ736112	BQ736112 AGENCOURT
30	895	69.8	758	13	BU320674	BU320674 603852801
31	892.5	69.6	1020	14	BY712345	BY712345 BY712345
32	891	69.4	912	12	BI915506	BI915506 603177449
33	875.5	68.2	933	12	BM475708	BM475708 AGENCOURT
34	874.5	68.2	657	12	BI850548	BI850548 imageqc_1
35	864.5	67.4	960	10	BF119347	BF119347 601757362
36	859.5	67.0	769	10	BF122628	BF122628 601760347
37	853.5	66.5	946	13	BQ720381	BQ720381 AGENCOURT
38	841	65.5	629	13	BU696861	BU696861 LL2in1295
39	839	65.4	685	14	BY734818	BY734818 BY734818
40	836.5	65.2	502	4	BX517485	Bx517485 RZPD Mus
41	836	65.2	2108	11	AK033543	AK033543 Mus muscu
42	833.5	65.0	1116	11	AK015705	AK015705 Mus muscu
43	831.5	64.8	1032	12	BM909967	BM909967 AGENCOURT
44	830.5	64.7	797	14	CA321457	CA321457 UI-M-FW0-
45	829.5	64.7	1108	13	BQ072716	BQ072716 AGENCOURT

ALIGNMENTS

RESULT 1
AL528502
LOCUS AL528502 1201 bp mRNA linear EST 23-MAY-2003
DEFINITION AL528502 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC029YJ10 5-PRIME, mRNA sequence.
ACCESSION AL528502
VERSION AL528502.2 GI:31066352
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)

```
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12791995.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC029DE05QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC029DE05QP1.
FEATURES              Location/Qualifiers
     source           1. .1201
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DC029YJ10"
                      /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      236 a    331 c    340 g    262 t     32 others
ORIGIN

Alignment Scores:
Pred. No.:                3.48e-107       Length:          1201
Score:                    1211.50         Matches:         226
Percent Similarity:       96.17%          Conservative:    0
Best Local Similarity:    96.17%          Mismatches:      7
Query Match:              94.43%          Indels:          2
DB:                       9               Gaps:            1

US-09-972-970-4 (1-233) x AL528502 (1-1201)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 231

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 291

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 351

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 411

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 471

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 531

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        532 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 591

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        592 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGASCCCGAGGC 651

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 711

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        712 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 771

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||||      |||||
Db        772 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 831

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        832 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 874

RESULT 2
AK008761
LOCUS       AK008761               1494 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male stomach cDNA, RIKEN full-length enriched
            library, clone:2210021G21 product:F-BOX PROTEIN FBX23 (FRAGMENT)
            homolog [Homo sapiens], full insert sequence.
ACCESSION   AK008761
VERSION     AK008761.1  GI:12843154
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
```

Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1494)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

FEATURES Location/Qualifiers
source 1. .1494
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2210021G21"
/db_xref="MGI:1907404"
/db_xref="taxon:10090"
/clone="2210021G21"
/sex="male"
/tissue_type="stomach"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 136. .948
/note="unnamed protein product; F-BOX PROTEIN FBX23 (FRAGMENT) homolog [Homo sapiens] (SPTR|Q9UKB9, evidence: FASTY, 86.9%ID, 99.1%length, match=363)
putative"
/codon_start=1
/protein_id="BAB25880.1"
/db_xref="GI:12843155"
/db_xref="MGI:1921507"
/translation="MPGKHQQFQDPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL IFFLELAAGILAFVFKDWIRDQLNLFINNNVKAYRDDLDLQNLIDFAQEYWSCCGARG PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDIRLKLELEQEGSI YTKGCVGQFEKWLQDNLIVVAGVLVGIALLQIFGLCLAQNLVSDIKAVKANW"
polyA_signal 1479. .1484
/note="putative"
polyA_site 1494
/note="putative"
BASE COUNT 257 a 414 c 463 g 360 t
ORIGIN

Alignment Scores:

Pred. No.:	5.55e-103	Length:	1494
Score:	1169.50	Matches:	220
Percent Similarity:	94.89%	Conservative:	3
Best Local Similarity:	93.62%	Mismatches:	10
Query Match:	91.15%	Indels:	2
DB:	11	Gaps:	1

US-09-972-970-4 (1-233) x AK008761 (1-1494)

```
Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||   ||||||:::||||||||||||||||||||||||||||||
Db        136 ATGCCCGGCAAGCACCAGCAATTCCAGGACCCTGAGGTCGGCTGCTGCGGGAAATACTTC 195

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATTGGCCTC 255

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCCAACATCTCGGCGCTGACAGATCTGGGCGGT 315

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CTTGACCCCGTGTGGCTGTTTGTGGTGGTGGGGGGAGTCATGTCAGTGTTGGGCTTCGCC 375

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GGCTGCATTGGGGCCCTCCGGGAAAACACCTTCCTGCTCAAGTTCTTCTCTGTGTTCCTC 435

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        436 GGCCTCATCTTCTTCCTGGAGCTGGCCGCGGGGATCCTGGCCTTCGTCTTCAAGGATTGG 495

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||:::
Db        496 ATCCGAGACCAACTTAATCTCTTCATCAACAACAATGTCAAAGCCTACCGGGATGATCTC 555

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 GACCTTCAGAACCTTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGTGCCCGAGGG 615

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCAAGCCGTGAG 675

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    |||||||||   ||
Db        676 CGCTGTGGGGTGCCCTTTCCTGCTGTGTTAGGGACCCAGCGGAAGATGTCCTCAATACC 735

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||||   |||  ||||||||||||  |||  |||||||||  |||||
Db        736 CAGTGTGGCTATGACATCCGACTCAAACTGGAGCTCGAGCAGGAGGGCTCCATCTACACC 795

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              ||||||||||||||||||||||||||||||  :::||||||
Db        796 AAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTACAAGACAACC 838
```

RESULT 3
AL529630
LOCUS AL529630 1194 bp mRNA linear EST 23-MAY-2003
DEFINITION AL529630 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens cDNA clone CS0DD005YB20 5-PRIME, mRNA sequence.
ACCESSION AL529630

```
VERSION     AL529630.2  GI:31067473
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12793123.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            3528.f For more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DD005DA10QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DD005DA10QP1.
FEATURES              Location/Qualifiers
     source           1. .1194
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DD005YB20"
                      /tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
                      /clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      237 a    331 c    336 g    253 t     37 others
ORIGIN

Alignment Scores:
Pred. No.:              1.59e-102       Length:         1194
Score:                  1163.50         Matches:        223
Percent Similarity:     95.32%          Conservative:   1
Best Local Similarity:  94.89%          Mismatches:     9
Query Match:            90.69%          Indels:         3
DB:                     9               Gaps:           1

US-09-972-970-4 (1-233) x AL529630 (1-1194)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 221

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 281

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db        282 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGAWCTGGGAGGC 341

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 401

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 461

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 521

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        522 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 581

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              :::||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        582 AACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGC-CGAGGC 640

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 700

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        701 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 760

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||  ||||||||||||      |||||
Db        761 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCWGGAGCAGCAGGGCTTCATCCACACC 820

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        821 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 863

RESULT 4
BY708665
LOCUS       BY708665                1005 bp    mRNA    linear   EST 16-DEC-2002
DEFINITION  BY708665 RIKEN full-length enriched, adult male stomach Mus
            musculus cDNA clone 2210021G21 5', mRNA sequence.
ACCESSION   BY708665
VERSION     BY708665.1  GI:27119857
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
```

```
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="2210021G21"
                     /sex="male"
                     /tissue_type="stomach"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /clone_lib="RIKEN full-length enriched, adult male
                     stomach"
                     /note="Site_1: XhoI; Site_2: SstI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and SstI."
BASE COUNT      175 a    281 c    310 g    237 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              3.87e-102       Length:         1005
Score:                  1158.50         Matches:        216
Percent Similarity:     93.59%          Conservative:   3
Best Local Similarity:  92.31%          Mismatches:     14
Query Match:            90.30%          Indels:         1
DB:                     14              Gaps:           1

US-09-972-970-4 (1-233) x BY708665 (1-1005)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||||||||||||||||||  ||||||:::|||||||||||||||||||||||||||||||
Db        136 ATGCCCGGCAAGCACCAGCAATTCCAGGACCCTGAGGTCGGCTGCTGCGGGAAATACTTC 195

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATTGGCCTC 255

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCCAACATCTCGGCGCTGACAGATCTGGGCGGT 315

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CTTGACCCCGTGTGGCTGTTTGTGGTGGTGGGGGGAGTCATGTCAGTGTTGGGCTTCGCC 375

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 CGCTGCATTGGGGCCCTCCGGGAAAACACCTTCCTGCTCAAGTTCTTCTCTGTGTTCCTC 435

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db        436 GGCCTCATCTTCTTCCTGGAGCTGGCCGCGGGGATCCTGGCCTTCGTCTTCAAGGATTGG 495

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||:::
Db        496 ATCCGAGACCAACTTAATCTCTTCATCAACAACAATGTCAAAGCCTACCGGGATGATCTC 555

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 GACCTTCAGAACCTTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGTGCCCGAGGG 615

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCAAGCCGTGAG 675

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspPro---AlaMetSerSerThrPro 199
              |||||||||||  |||||||||||||||||||||||||         |||||||||   |||
Db        676 CGCTGTGGGGTCGCCTTTTCCTGCTGTGTTAGGGACCAGCGGAAGATGTCCTCAATACCC 735

Qy        200 SerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPro 219
              |||||||||||  |||  ||||||||||||  |||  |||||||||   ||||||
Db        736 AGTGTGGCTATGACATCCGACTCAAACTGGAGCTCGAGCAGGAGGGCTCCATCTACACCA 795

Qy        220 LysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||||||||||||||||||||||||||   :::||||||
Db        796 AAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTACAAGACAACC 837

RESULT 5
BE615772
LOCUS       BE615772                 849 bp    mRNA    linear   EST 24-AUG-2000
DEFINITION  601279927F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3621871 5',
            mRNA sequence.
ACCESSION   BE615772
VERSION     BE615772.1  GI:9897371
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM296  row: e  column: 08
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .849
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3621871"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_39"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California,  Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies)."
BASE COUNT      165 a    241 c    243 g    199 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.49e-101      Length:          849
Score:                  1151.50        Matches:         218
Percent Similarity:     93.59%         Conservative:    1
Best Local Similarity:  93.16%         Mismatches:      14
Query Match:            89.75%         Indels:          2
DB:                     10             Gaps:            1

US-09-972-970-4 (1-233) x BE615772 (1-849)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         33 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAANC-TTC 91

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         92 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 151

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        152 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 211

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        212 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 271

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        272 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 331

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        332 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 391

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        392 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 451

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 511

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 571

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAlaMetSer---SerThrPro 199
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||||
Db        572 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAAACCC 631

Qy        200 SerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPro 219
              |||||||||      ||||||||||||||||||||||||||||||||      |||||||||
Db        632 AGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACAA 691

Qy        220 LysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              ||||||   ||||||||||||||||||||||   :::   |||
Db        692 AAGGCTGCGTGGGCCAGTCTGAGAAGTGGCTGGCAGGAAACC 733

RESULT 6
BU856801
LOCUS       BU856801                 938 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_10475185 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6646576 5', mRNA sequence.
ACCESSION   BU856801
VERSION     BU856801.1  GI:24041791
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2885  row: j  column: 16
            High quality sequence stop: 649.
FEATURES             Location/Qualifiers
     source          1. .938
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6646576"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_107"
                     /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      162 a    290 c    279 g    207 t
ORIGIN

Alignment Scores:
Pred. No.:              1.89e-101      Length:          938
Score:                  1151.00        Matches:         218
Percent Similarity:     93.22%         Conservative:    2
Best Local Similarity:  92.37%         Mismatches:      13
Query Match:            89.71%         Indels:          3
DB:                     13             Gaps:            2

US-09-972-970-4 (1-233) x BU856801 (1-938)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        113 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 172

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 232

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 292

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        293 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 352

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 412

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 472

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        473 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 532

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 592

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 652

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    ||||||    |||||
Db        653 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCCTCACACC 712

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      |||||||||||||||||||||||||:::|||      |||||
Db        713 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAAGGCTTCATCCACACC 772

Qy        219 oLysAlaValTrpAlaSerLeu---ArgSerGlyCysArgThrThr 233
              |||||||      |||         :::         |||||||||
Db        773 AAAGGCTGCGGGGGCCCAATTTGAAAAAGTGGCTGCAAGGACCACC 818

RESULT 7
CA454987
LOCUS       CA454987                901 bp    mRNA    linear   EST 12-NOV-2002
DEFINITION  AGENCOURT_10735735 MAPcL Homo sapiens cDNA clone IMAGE:6722499 5',
            mRNA sequence.
ACCESSION   CA454987
VERSION     CA454987.1  GI:24905277
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 901)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Kristi A. Egland, Ira Pastan
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14285  row: f  column: 03
            High quality sequence stop: 683.
FEATURES             Location/Qualifiers
     source          1. .901
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6722499"
                     /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HME1
                     , LNCaP"
                     /lab_host="EMDH10B"
                     /clone_lib="MAPcL"
                     /note="Vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
                     Subtracted with brain, liver, lung, kidney and muscle.
                     Directionally cloned. Priming method: oligo-dT. Average
                     insert size: 1800 bp. Library amplification:  26,000 fold.
                     Kristi A. Egland, James J. Vincent, Robert Strausberg,
                     Bungkook Lee  & Ira Pastan:  Discovery of new breast
                     cancer genes encoding  membrane and secreted proteins.
                     Manuscript submitted."
BASE COUNT      157 a    275 c    269 g    200 t
ORIGIN

Alignment Scores:
Pred. No.:              1.35e-95        Length:         901
Score:                  1090.50         Matches:        206
Percent Similarity:     93.69%          Conservative:   2
Best Local Similarity:  92.79%          Mismatches:     11
Query Match:            85.00%          Indels:         3
DB:                     14              Gaps:           1

US-09-972-970-4 (1-233) x CA454987 (1-901)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 192

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 252

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 312

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 372

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 432

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 492

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||  ||||||||||||||||||||||
Db        493 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGACAAGGCCTACCGGGACGACATT 552

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 612

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 672

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    ||||||    |||||
Db        673 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACAAC 732

Qy        199 oSerValAlaMetMetSer-GlySerAsnTrpSerTrpSerSerArgAlaProTyrThrP 219
              |||||||||      |||  ||||||||||||||      :::      :::   |
Db        733 CAGTGTGGCTACAACGTCCGAGCTCCAACTGGAGCTGGGAACAACAAGGCCTTCATCCCC 792

Qy        219 ro 219
              ||
Db        793 CA 794

RESULT 8
BG770931
LOCUS       BG770931                781 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602719148F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4839779 5',
            mRNA sequence.
ACCESSION   BG770931
```

```
VERSION     BG770931.1  GI:14081584
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1669  row: o  column: 12
            High quality sequence stop: 697.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4839779"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_60"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      127 a    239 c    239 g    176 t
ORIGIN

Alignment Scores:
Pred. No.:              2.19e-95        Length:         781
Score:                  1087.50         Matches:        203
Percent Similarity:     94.88%          Conservative:   1
Best Local Similarity:  94.42%          Mismatches:     10
Query Match:            84.76%          Indels:         2
DB:                     12              Gaps:           1

US-09-972-970-4 (1-233) x BG770931 (1-781)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 188

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 248

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 308

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 368

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        369 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 428

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 488

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        489 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 548

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 608

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 668

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSer---ThrPro 199
              ||||||||||||||||||||||||||||||:::|||||||||                |||
Db        669 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCCAAGACCCTGCGGAAGGATGTCTCAAACCC 728

Qy        200 SerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArg 214
              |||||||||      ||||||   ||||||||||||      |||
Db        729 AGTGTGGCTACGACGTCCGGTCA-AACTGGAGCTGGAGGCAGAGG 772

RESULT 9
BM563474
LOCUS       BM563474                1032 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6589677 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441687
            5', mRNA sequence.
ACCESSION   BM563474
VERSION     BM563474.1  GI:18810427
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1916  row: f  column: 24
            High quality sequence stop: 697.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5441687"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      203 a    309 c    286 g    232 t      2 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:                8.62e-95        Length:         1032
Score:                    1083.00         Matches:        204
Percent Similarity:       86.81%          Conservative:   0
Best Local Similarity:    86.81%          Mismatches:     6
Query Match:              84.41%          Indels:         25
DB:                       12              Gaps:           2

US-09-972-970-4 (1-233) x BM563474 (1-1032)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 115

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 175

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 235

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 295

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||
Db        296 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTT---------------- 340

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
                                                                  ||||||
Db        341 ------------------------------------------------------GACTGG 346

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        347 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 406

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 466

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 526

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        527 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 586

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||||      |||||
Db        587 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 646

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        647 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 689

RESULT 10
AL543914
LOCUS       AL543914                957 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL543914 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI005YM07 5-PRIME, mRNA sequence.
ACCESSION   AL543914
VERSION     AL543914.2  GI:31265759
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 957)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12876393.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI005AG04QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI005AG04QP1.
FEATURES              Location/Qualifiers
     source           1. .957
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI005YM07"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      176 a    279 c    288 g    205 t      9 others
ORIGIN

Alignment Scores:
Pred. No.:                1.71e-94        Length:         957
Score:                    1079.50         Matches:        208
Percent Similarity:       88.19%          Conservative:   1
Best Local Similarity:    87.76%          Mismatches:     12
Query Match:              84.14%          Indels:         16
DB:                       9               Gaps:           2

US-09-972-970-4 (1-233) x AL543914 (1-957)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 231

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 291

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 351

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 411

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||
Db        412 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTT---------------- 456

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
                                  ||||||||||||||||||||||||||||||||||||||||||||
Db        457 --------------------CTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 495

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        496 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 555

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGln--GluTyrTrpSerCysCysGlyAlaArgG 160
              ||||||||||||||||||||||||||||        :::||||||||||||||||
Db        556 GACCTCCAGAACCTCATTGACTTTGCTCATGTGTWTTCTGGTCTTGCTGCGGASCCGAGG 615

Qy        160 lyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgG 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        616 CCCTCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGG 675

Qy        180 luArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThr 198
              ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        676 AGCGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACA 735

Qy        199 ProSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThr 218
              |||||||||||||      ||||||||||||||||||||||||||||||||      |||
Db        736 CCCAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACA 795

Qy        219 ProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||||   ||||||||||||||||||||||||||||||||
Db        796 CCAAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 840

RESULT 11
AL556793
LOCUS       AL556793                1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL556793 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DK005YH21 5-PRIME, mRNA sequence.
ACCESSION   AL556793
VERSION     AL556793.2  GI:31278594
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12899797.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DK005CD11QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DK005CD11QP1.
FEATURES              Location/Qualifiers
     source           1. .1201
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DK005YH21"
                      /cell_type="HELA CELLS COT 25-NORMALIZED"
                      /cell_line="HELA"
                      /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      231 a    322 c    324 g    258 t     66 others
ORIGIN

Alignment Scores:
Pred. No.:              2.87e-94        Length:         1201
Score:                  1078.50         Matches:        209
Percent Similarity:     92.54%          Conservative:   2
Best Local Similarity:  91.67%          Mismatches:     11
Query Match:            84.06%          Indels:         7
DB:                     9               Gaps:           3

US-09-972-970-4 (1-233) x AL556793 (1-1201)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 327

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 387

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 447

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        448 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 507

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        508 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 567

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        568 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 627

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        628 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 687

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        688 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGASC-CGAGGC 746

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 806

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        807 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATKTCCTCAACACC 866

Qy        199 oSerValAlaMetMetSerGlySerAsnTrp---------SerTrpSerSerArgAlaPr 216
              ||||||||||      ||||||||||||||         |||      :::
Db        867 CAGTGTGGCTACGACGTCCGGCTCAAACTGGTGAGAGGGGAGCTKGASAGCAGGSYTYAT 926

Qy        216 oTyrThrProLysAlaValTrp 223
              :::|||   ||||||   |||
Db        927 YCACACA---AAAGCTGCGTGG 945

RESULT 12
BG477727
LOCUS       BG477727                1028 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602521422F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4639997 5',
            mRNA sequence.
ACCESSION   BG477727
VERSION     BG477727.1  GI:13410006
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1028)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1404  row: k  column: 06
            High quality sequence stop: 734.
FEATURES              Location/Qualifiers
     source           1. .1028
```

```
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4639997"
                  /tissue_type="melanotic melanoma"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH_MGC_20"
                  /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                  EcoRI; cDNA made by oligo-dT priming.  Directionally
                  cloned into EcoRI/XhoI sites using the following 5'
                  adaptor: GGCACGAG(G).  Size-selected >500bp for average
                  insert size 1.8kb.  Library constructed by Ling Hong in
                  the laboratory of Gerald M. Rubin (University of
                  California, Berkeley) using  ZAP-cDNA synthesis kit
                  (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      194 a    296 c    310 g    228 t
ORIGIN

Alignment Scores:
Pred. No.:               1.12e-93       Length:         1028
Score:                   1071.50        Matches:        214
Percent Similarity:      91.10%         Conservative:   1
Best Local Similarity:   90.68%         Mismatches:     17
Query Match:             83.52%         Indels:         6
DB:                      10             Gaps:           1

US-09-972-970-4 (1-233) x BG477727 (1-1028)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 193

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 253

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 313

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 373

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 433

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 493

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        494 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGCGACGACATT 553

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db        554 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGG- 612

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsn-CysThrAspLeuAsnProSerArgGl 180
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        613 CCCAATGACTGGAACCTCAATATCTACTTCAACTTGCACTGACTTGAACCCCAGCCGGGA 672

Qy        180 uArgCysGlyValProPheSerCysCysValArgAsp---ProAlaMetSerSerThrPr 199
              |||||||||||    ||||||||||||||||||                ||||||||   ||
Db        673 GCGCTGCGGGGTGGCCTTCTCCTGCTGCGTAAGGACCTGGCGGAGGATGTCCTCAAAACC 732

Qy        199 oSerVal-AlaMetMetSerGlySerAsnTrp-SerTrpSerSerArgAlaProTyrThr 218
              |||||| |||         ||||||||||||| ||||||||||||||           |||
Db        733 CAGTGTGGGCTACAACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGGTTCATCCACA 792

Qy        219 ProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              ||||||    |||||||||   :::     |||||||||
Db        793 CCAAAGGT-GCGTGGCCAGTTGAAAAGTGGCTTGCAGGACA 833

RESULT 13
BI914325
LOCUS       BI914325                 718 bp    mRNA    linear   EST 17-OCT-2001
DEFINITION  603182305F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246304 5',
            mRNA sequence.
ACCESSION   BI914325
VERSION     BI914325.1  GI:16198834
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11621  row: j  column: 01
            High quality sequence stop: 712.
FEATURES             Location/Qualifiers
     source          1..718
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5246304"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_121"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
                     fetal brains, female age 20 weeks, female age 24 weeks,
                     and male age 26 weeks.  Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     0.7-3.5 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 017. Note:
                     this is a NIH_MGC Library."
BASE COUNT      115 a    222 c    217 g    164 t
ORIGIN

Alignment Scores:
Pred. No.:               2.4e-93        Length:         718
Score:                   1066.00        Matches:        193
Percent Similarity:      99.48%         Conservative:   0
Best Local Similarity:   99.48%         Mismatches:     1
Query Match:             83.09%         Indels:         0
DB:                      12             Gaps:           0

US-09-972-970-4 (1-233) x BI914325 (1-718)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 178

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 238

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 298
```

```
Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 358

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 418

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 478

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        479 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 538

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 598

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 658

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla 194
              |||||||||||||||||||||||||||||||||||||||||
Db        659 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCG 700
```

```
RESULT 14
BG017161
LOCUS       BG017161                714 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  de61f11.y1 Kirschner embryo St10 14 Xenopus laevis cDNA clone
            IMAGE:3516428 5' similar to TR:O60628 O60628 TETRASPANIN TSPAN-5.
            [1] ;, mRNA sequence.
ACCESSION   BG017161
VERSION     BG017161.1  GI:12471136
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished
COMMENT     Other_ESTs: de61f11.x1
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by M. Kirschner (Harvard Medical School). DNA
            Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 442.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3516428"
                     /tissue_type="pooled embryos (stage 10-14)"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="Kirschner embryo St10 14"
                     /note="Vector: pCS2+; Site_1: NotI; Site_2: SalI;
                     Size-selected for average insert size 1.2 kb. Library was
                     constructed and donated by M. Kirschner (Harvard Medical
                     School)."
BASE COUNT      170 a    154 c    177 g    212 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                2.56e-89        Length:          714
Score:                    1024.50         Matches:         190
Percent Similarity:       85.84%          Conservative:    10
Best Local Similarity:    81.55%          Mismatches:      31
Query Match:              79.85%          Indels:          2
DB:                       10              Gaps:            1

US-09-972-970-4 (1-233) x BG017161 (1-714)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   |||||||||||||||||:::   |||||||||   ||||||   |||||||||
Db         10 ATGTCAGGCAAACATCAGCATTTTAAAGGTCCCGAAGTGAGCTGCTGTGTGAAGTATTTC 69

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||:::|||||||||   ||||||||||||||||
Db         70 TTGTTTGGATTCAACATAGTGTTTTGGTTGCTTGGAGCTAGTTTTCTAGCTATTGGCTTT 129

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||   ||||||||||||||||||||||   ::::::||||||||||||||||
Db        130 TGGGCATGGGCAGAAAAGGGTGTCCTGTCCAACATCTACTCCATCACAGACCTTGGTGGC 189

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||:::|||   |||   ||||||   ||||||||||||||||
Db        190 TTTGACCCCGTTTGGCTCTTCATAGTCACTGGAGCTGTCATGTTTGTGTTGGGGTTTGCC 249

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        250 GGCTGCATTGGAGCCTTGAGAGAAAACACGACACTCCTAAAATTTTTCTCAGTCTTCCTG 309

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        310 GGGTTGATCTTCTTCTTGGAGCTGACTGCAGGGATCTTGGCCTTTGTTTTCAAAGACTGG 369

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 ATCAAAGATCAGCTTAATTTTTTTTATCAACAACAACGTGAAGGCGTATCGTGATGATATT 429

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        430 GACCTCCAGAACCTCATTGATTTTGCTCAGGAATATTGGTCCTGCTGTGGTGCACATGGA 489

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db        490 CCTAACGACTGGAACCTCAACATCTACTTTAACTGTACAGACTCTAACCCAAGCAGGGAA 549

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||   |||||||||:::|||||||||    |||       |||
Db        550 CGCTGCGGTGTGCCATTCTNCTGCTGTGTGAAGGACCCTGCGGAAGATGTTCCCAACACA 609

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||   ||||||   |||   :::|||:::|||||||||   |||          ||
Db        610 CAGTGTGTATATGATGTGCGGCTGAAGCTGGAACTGGAGCAGCATAGCTTCATCTATACC 669

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArg 231
              ||||   |||||||||   |||   |||||||||:::
Db        670 AAAGGATGTGTGGGCCCAGTTGAGCAGTGGCTGCAAG 706
```

```
RESULT 15
BG478644
LOCUS     .   BG478644                902 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602525532F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4643959 5',
            mRNA sequence.
ACCESSION   BG478644
VERSION     BG478644.1  GI:13410923
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 902)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1414  row: p  column: 08
            High quality sequence stop: 813.
FEATURES             Location/Qualifiers
     source          1. .902
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4643959"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_20"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      146 a    268 c    288 g    200 t
ORIGIN

Alignment Scores:
Pred. No.:              4.85e-89        Length:         902
Score:                  1023.00         Matches:        210
Percent Similarity:     88.61%          Conservative:   0
Best Local Similarity:  88.61%          Mismatches:     19
Query Match:            79.73%          Indels:         9
DB:                     10              Gaps:           1

US-09-972-970-4 (1-233) x BG478644 (1-902)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 195

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 255

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAsp-LeuGlyGl 60
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        256 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCCTGGGAGG 315

Qy         60 yLeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAl 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CCTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGC 375

Qy         80 aGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLe 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 TGGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCT 435

Qy        100 uGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 CGGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTG 495

Qy        120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAsp-AspI 140
              ||||||||||||||||  |||||||||||||||||||||||||||||||||| |||| |
Db        496 GATTCGAGACCAGCTCAAACTCTTCATCAACAACAACGTCAAGGCCTACCGGGACCGACA 555

Qy        140 leAspLeuGlnAsnLeuIleAspPheAla-GlnGluTyrTrpSerCysCysGlyAlaArg 159
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        556 TTGACCTCCAGAACCTCATTGACTTTGCCTCAGGAATACTGGTCTTGCTGTGGAGCCCGA 615

Qy        160 GlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCys-ThrAspLeuAsnProSerAr 179
              |||||||||||||||||||||||||||||||||||||| |||||||||   ||||||||
Db        616 GG-CCCAATGACTGGAACCTCAATATCTACTTCAACTGCAACTGACTTGAGACCCAGCCG 674

Qy        179 gGluArgCys-GlyValProPheSerCysCysValArgAspProAla------MetSerS 197
              |||||||||| ||||||||||||||||||||||||||||||||||||||          |
Db        675 GGAGCGCTGCAGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGGAGGATGTCCCT 734

Qy        197 erThrProSerVal-AlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaPro 216
              ||      |||||| |||                |||||||||||||||||||||||
Db        735 CGAGAGCCAGTGTTGGCTACCGACGTCCGGTTCAACTGGAGCTGGAGCAGCAGGGCTTCA 794

Qy        217 TyrThrProLysAlaValTrpAlaSerLeuArgSerGly 229
                      ||||||   |||||||||   |||||||||
Db        795 TCCCAGACAAAGGCTGCGTGGGCCAGTTGGAGAAGTGGT 833
```

Search completed: November 21, 2003, 17:13:02
Job time : 2160 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 13:30:43 ; Search time 280 Seconds
(without alignments)
2246.318 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972970/runat_21112003_125335_27199/app_query.fasta_1.391
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972970 @CGN_1_1_312_@runat_21112003_125335_27199 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1283	100.0	2538	24	AAD21883	Human TM4SF recept
2	1215.5	94.7	2091	21	AAA96482	cDNA encoding a hu
3	1215.5	94.7	3185	24	AAI72633	Human 23228 cDNA.
4	1212.5	94.5	1126	22	AAC90014	Clone HNTMH27 codi
5	1204.5	93.9	1644	24	ABL90838	Human polynucleoti
6	1176.5	91.7	1023	24	ABZ11816	Human polynucleoti
7	958.5	74.7	1655	21	AAC98195	Human colon cancer
8	955.5	74.5	1567	23	ACC44092	Rat gamma-hydroxyb
9	918.5	71.6	1174	22	AAC90015	Clone HE8EJ16 codi
10	892.5	69.6	1178	22	AAC90020	Clone HE8EL16 codi
11	807	62.9	1110	20	AAZ20854	Polynucleotide seq
12	807	62.9	1110	22	AAS59291	Human cDNA encodin
13	807	62.9	1110	24	ABA90960	Human polynucleoti
14	790	61.6	864	24	AAI72287	NET-4 antisense mo
15	759.5	59.2	1988	24	ABK35735	cDNA sequence #126
16	654.5	51.0	813	22	AAF90629	Human TANGO 339 cD
17	654.5	51.0	1388	24	AAD29623	Human tetraspan pr
18	654.5	51.0	2465	25	ABX70953	Novel human cDNA s
19	654.5	51.0	2715	22	AAF90628	Human TANGO 339 cD
20	653.5	50.9	813	22	AAF90656	Human TANGO 339 K1
21	651.5	50.8	813	22	AAF90657	Human TANGO 339 F2
22	651.5	50.8	2623	23	AAS92016	DNA encoding novel
23	650.5	50.7	813	22	AAF90658	Human TANGO 339 A4
24	650.5	50.7	813	22	AAF90659	Human TANGO 339 D6
25	641.5	50.0	2672	21	AAZ65259	Human secreted pro
26	635.5	49.5	1428	21	AAA93623	Human tetraspanin-
27	583.5	45.5	1667	21	AAZ65350	Human secreted pro
28	556	43.3	601	23	AAS80667	DNA encoding novel
29	537	41.9	368	21	AAZ93364	Sequence encoding
30	537	41.9	368	24	AAL41055	cDNA of Human F-bo
31	490.5	38.2	839	22	AAH34927	Human colon cancer
32	482.5	37.6	810	23	ABL14835	Drosophila melanog
33	476.5	37.1	852	22	AAS44553	Human cDNA encodin
34	476.5	37.1	1932	21	AAZ98047	Human secreted pro
35	476.5	37.1	1932	22	AAD11660	Human secreted pro
36	476.5	37.1	1932	24	ABK69756	Human secreted pro
37	476.5	37.1	2384	24	ABZ11836	Human polynucleoti
38	461.5	36.0	2638	24	ABL90058	Human polynucleoti
39	441	34.4	639	21	AAA60182	Hydrophobic domain
40	441	34.4	2367	21	AAA60192	Hydrophobic domain
41	430	33.5	3009	23	ABL14834	Drosophila melanog
42	427.5	33.3	816	21	AAA93675	Human tetraspanin-
43	421.5	32.9	1187	23	ABL07929	Drosophila melanog
44	411.5	32.1	1120	23	ABL16799	Drosophila melanog
c 45	402	31.3	3244	23	ABL07928	Drosophila melanog

ALIGNMENTS

RESULT 1
AAD21883
ID AAD21883 standard; cDNA; 2538 BP.
XX
AC AAD21883;
XX
DT 12-FEB-2002 (first entry)
XX
DE Human TM4SF receptor-encoding gene 1 cDNA clone HOFOB55, SEQ ID NO:2.
XX
KW Human; 4-transmembrane superfamily; TM4SF receptor; breast cancer;
KW ovarian cancer; immune disorder; Addison's disease; wound healing;
KW gene therapy; autoimmune haemolytic anaemia; autoimmune thyroiditis;
KW diabetes mellitus; allergy; Crohn's disease; multiple sclerosis;
KW rheumatoid arthritis; ulcerative colitis; cardiovascular disorder;
KW myocardial ischaemia; neurological disease; antifungal; antiviral;
KW antibacterial; cerebral anoxia; epilepsy; infectious disease;

```
KW   antiparasitic; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             119..820
FT                   /*tag=  a
FT                   /product= "Human TM4SF receptor protein"
XX
PN   WO200177173-A1.
XX
PD   18-OCT-2001.
XX
PF   05-APR-2001; 2001WO-US11130.
XX
PR   10-APR-2000; 2000US-195336P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Shi Y,  Ruben SM;
XX
DR   WPI; 2002-017447/02.
DR   P-PSDB; AAE13218.
XX
PT   Novel isolated protein, a member of 4-transmembrane superfamily of
PT   receptor polypeptides, useful in the prevention, treatment and
PT   diagnosis of cancer, immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 1; Page 265-266; 271pp; English.
XX
CC   The invention relates to human 4-transmembrane superfamily (TM4SF)
CC   receptor polypeptides and polynucleotides. Sequences of the invention
CC   are useful for preventing, treating, ameliorating or diagnosing a
CC   pathological condition or a susceptibility to a pathological condition.
CC   TM4SF polypeptides are useful for screening molecules which modify
CC   their activity. TM4SF nucleic acids, protein, antibodies, agonists and
CC   antagonists are useful in the diagnosis, treatment and prevention of
CC   cancer, particularly breast and ovarian cancer, and other cancers of
CC   the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC   liver, lung, or urogenital; immune disorders such as Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC   bacterial, fungal and parasitic infections. TM4SF polynucleotides are
CC   also useful in gene therapy. The present sequence is human TM4SF
CC   receptor protein encoding cDNA.
XX
SQ   Sequence 2538 BP; 528 A; 669 C; 747 G; 594 T; 0 other;

Alignment Scores:
Pred. No.:              1.59e-114      Length:          2538
Score:                  1283.00        Matches:         233
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     24             Gaps:            0

US-09-972-970-4 (1-233) x AAD21883 (1-2538)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTC 178

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 CTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTC 238

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGC 298

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCC 358

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 GGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTT 418

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGG 478

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 ATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATT 538

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGG 598

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAG 658

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrProSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 CGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTCAACACCCAGT 718

Qy        201 ValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778

Qy        221 AlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||||||||||||||||||||||||||||||||||
Db        779 GCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 817

RESULT 2
AAA96482
ID   AAA96482 standard; cDNA; 2091 BP.
XX
AC   AAA96482;
XX
DT   08-FEB-2001  (first entry)
XX
DE   cDNA encoding a human transmembrane protein.
XX
KW   Human; transmembrane protein; cell proliferation disorder;  myeloma;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW   allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW   Alzheimer's disease; Tourette's disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             113..1102
FT                   /*tag=  a
XX
PN   WO200056891-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07817.
XX
PR   22-MAR-1999;   99US-0125537.
PR   16-JUN-1999;   99US-0139565.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Yue H,  Lal P,  Tang YT,  Hillman JL,  Reddy R,  Bandman O,  Baughn MR;
PI   Lu DAM,  Azimzai Y,  Yang J;
XX
```

```
DR   WPI; 2000-579485/54.
DR   P-PSDB; AAB18969.
XX
PT   New human transmembrane proteins are used to treat a disease or
PT   condition associated with decreased expression of functional HTMP e.g.
PT   Tourette's disorder, angina and leukaemia -
XX
PS   Claim 4; Page 113-114; 130pp; English.
XX
CC   The present sequence encodes a human transmembrane proteins (HTMP).
CC   Agonists and antagonists of the protein are used to treat a disease
CC   or condition associated with overexpression of the protein. Diseases
CC   and conditions which can be treated include cell proliferative,
CC   immunological, reproductive, smooth muscle and neurological disorders
CC   e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC   syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC   stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC   polynucleotides may be used to detect and quantify gene expression in
CC   biopsied tissues where protein expression may be correlated with disease
CC   e.g. to determine absence, presence or excess expression of HTMP or to
CC   monitor regulation of HTMP expression during therapeutic intervention.
XX
SQ   Sequence 2091 BP; 392 A; 602 C; 637 G; 460 T; 0 other;

Alignment Scores:
Pred. No.:                4.4e-108        Length:          2091
Score:                    1215.50         Matches:         227
Percent Similarity:       96.60%          Conservative:    0
Best Local Similarity:    96.60%          Mismatches:      6
Query Match:              94.74%          Indels:          2
DB:                       21              Gaps:            1

US-09-972-970-4 (1-233) x AAA96482 (1-2091)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        113 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 172

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 232

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 292

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        293 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 352

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 412

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 472

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        473 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 532

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 592

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 652

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        653 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 712

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||        ||||||||||||||||||||||||||||||||      |||||
Db        713 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 772

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        773 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 815

RESULT 3
AAI72633
ID   AAI72633 standard; cDNA; 3185 BP.
XX
AC   AAI72633;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human 23228 cDNA.
XX
KW   Gene; tetraspanin; 23228; cell surface protein; transmembrane domain;
KW   extracellular growth factor; HB-EGF; TGF-alpha; amphiregulin;
KW   diabetes mellitus; arthritis; multiple sclerosis; encephalomyelitis;
KW   dermatitis; Crohn's disease; and asthma; cancer; metastasis;
KW   viral infection; cellular signalling activity; cell proliferation;
KW   cell motility; CD81; B-Cell antigen receptor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             168..980
FT                   /*tag=  a
FT                   /product= "23228"
XX
PN   WO200216603-A2.
XX
PD   28-FEB-2002.
XX
PF   21-AUG-2001; 2001WO-US41811.
XX
PR   21-AUG-2000; 2000US-226612P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Leiby KR;
XX
DR   WPI; 2002-315466/35.
DR   P-PSDB; AAB47950.
XX
PT   New tetraspanin 23228 polypeptide useful in screening assays,
PT   predictive medicine and as a prophylactic or therapeutic agent, e.g.,
PT   for hematopoietic and immune diseases such as diabetes or multiple
PT   sclerosis -
XX
PS   Claim 1; Page 111-113; 123pp; English.
XX
CC   This sequence encodes tetraspanin 23228. This protein is a cell surface
CC   protein having four transmembrane domains (TM1-TM4).  TM1, TM3 and TM4
CC   have a single polar amino acid located within them, which may interact
CC   with each other and contribute to domain stability.  The cytoplasmic N-
CC   and C-terminal domains and the intracellular loop between TM2 and TM3
CC   are less than 30 amino acids in length.  The second extracellular domain
CC   between TM3 and TM4 contains conserved Cys residues and may function
CC   to bind extracellular growth factors, such as HB-EGF, TGF-alpha and
CC   amphiregulin. The 23228 protein is useful for diagnosing and treating
CC   23228-mediated disorders, e.g., haematopoietic and/or immune disorders
CC   such as diabetes mellitus, arthritis, multiple sclerosis,
CC   encephalomyelitis, dermatitis, Crohn's disease and asthma; cell
CC   proliferation and differentiation disorders, e.g., cancers or
CC   metastasis; and/or viral infections. 23228 polypeptide can control
CC   cellular signalling activity, bind to an extracellular growth factor,
CC   for example, amphiregulin, regulate cell proliferation, bind to a cell
CC   surface protein, to recruit intracellular kinases, to regulate cell
CC   motility, bind to another tetraspanin such as CD81, to associate with a
```

```
CC   B-Cell antigen receptor and the ability to modulate the association
CC   with a virus.
XX
SQ   Sequence 3185 BP; 676 A; 883 C; 945 G; 681 T; 0 other;

Alignment Scores:
Pred. No.:              7.7e-108        Length:         3185
Score:                  1215.50         Matches:        227
Percent Similarity:     96.60%          Conservative:   0
Best Local Similarity:  96.60%          Mismatches:     6
Query Match:            94.74%          Indels:         2
DB:                     24              Gaps:           1

US-09-972-970-4 (1-233) x AAI72633 (1-3185)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 227

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 287

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 347

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 407

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 467

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 527

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        528 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 587

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 647

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 707

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        708 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 767

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||||      |||||
Db        768 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 827

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        828 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 870

RESULT 4
AAC90014
ID   AAC90014 standard; cDNA; 1126 BP.
XX
AC   AAC90014;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Clone HNTMH27 coding sequence.
XX
KW   Gene therapy; human; 4 transmembrane superfamily receptor protein;
KW   endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW   reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW   bacterial infection; viral; fungal; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070076-A1.
XX
PD   23-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13504.
XX
PR   19-MAY-1999;   99US-0135122.
PR   03-JUN-1999;   99US-0137797.
PR   11-JUN-1999;   99US-0138573.
PR   18-AUG-1999;   99US-0149447.
PR   28-JAN-2000; 2000US-0178770.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Fan P,  Roschke V,  Shi Y,  Komatsoulis GA;
PI   Rosen CA;
XX
DR   WPI; 2001-007502/01.
DR   P-PSDB; AAB49504.
XX
PT   Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT   superfamily receptor protein, useful for diagnosing, treating and/or
PT   preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS   Claim 1; Pages 269-270; 297pp; English.
XX
CC   The present invention relates to isolated nucleic acids and proteins
CC   encoding human soluble 4 transmembrane superfamily receptor protein (see
CC   AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
CC   nucleic acid. The present sequence is useful for preventing, treating or
CC   ameliorating a medical condition and in diagnosing (susceptibility to) a
CC   pathological condition e.g. endocrine disorders e.g. Addison's disease,
CC   (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,
CC   cerebrovascular diseases, neural disorders e.g. Alzheimer's and
CC   Parkinson's disease, reproductive disorders, skin disorders e.g.
CC   psoriasis, renal system disorders e.g. nephritis, (auto)immune system
CC   disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.
CC   neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections
CC   caused by bacteria, viruses and fungi.
XX
SQ   Sequence 1126 BP; 191 A; 352 C; 343 G; 240 T; 0 other;

Alignment Scores:
Pred. No.:              3.77e-108       Length:         1126
Score:                  1212.50         Matches:        226
Percent Similarity:     96.60%          Conservative:   1
Best Local Similarity:  96.17%          Mismatches:     6
Query Match:            94.51%          Indels:         2
DB:                     22              Gaps:           1

US-09-972-970-4 (1-233) x AAC90014 (1-1126)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 216

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 276

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 336

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        337 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 396

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 456

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 516

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        517 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 576

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        577 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 636

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 696

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        697 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 756

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||      |||||
Db        757 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 816

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||:::||||||||||||||
Db        817 AAAGGCTGCGTGGGCCAGTTTGAGAAATGGCTGCAGGACAACC 859
```

```
RESULT 5
ABL90838
ID   ABL90838 standard; cDNA; 1644 BP.
XX
AC   ABL90838;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1400.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US16450.
XX
PR   19-MAY-2000; 2000US-205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   P-PSDB; ABB90429.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 4; SEQ ID NO 1400; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1644 BP; 297 A; 479 C; 513 G; 346 T; 9 other;
```

Alignment Scores:			
Pred. No.:	3.73e-107	Length:	1644
Score:	1204.50	Matches:	225
Percent Similarity:	95.74%	Conservative:	0
Best Local Similarity:	95.74%	Mismatches:	8
Query Match:	93.88%	Indels:	2
DB:	24	Gaps:	1

US-09-972-970-4 (1-233) x ABL90838 (1-1644)

```
Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 281

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 341

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        342 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTYAGCGCTGACAGATCTGGGAGGC 401

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 CTTGAMCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 461

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 521

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 581

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        582 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 641

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 701

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 761

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        762 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 821

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
```

```
              ||||||||||      ||||||||||||||||||||||||||||||||      |||||
Db        822 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 881

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        882 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 924

RESULT 6
ABZ11816
ID   ABZ11816 standard; cDNA; 1023 BP.
XX
AC   ABZ11816;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 698.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP69599.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders
XX
PS   Claim 1; SEQ ID NO 698; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1023 BP; 190 A; 300 C; 293 G; 240 T; 0 other;

Alignment Scores:

Pred. No.:                1.04e-104      Length:          1023
Score:                    1176.50        Matches:         227
Percent Similarity:       95.38%         Conservative:    0
Best Local Similarity:    95.38%         Mismatches:      6
Query Match:              91.70%         Indels:          5
DB:                       24             Gaps:            1

US-09-972-970-4 (1-233) x ABZ11816 (1-1023)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 90

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 150

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 210

Qy         61 LeuAspProValTrp-LeuPheValValValGlyGlyValMetSerValLeuGlyPheAl 80
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        211 CTTGACCCCGTGTGGCTTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGC 270

Qy         80 a-GlyCys-IleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPhe 99
              | |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 TGGGCTGCAATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTC 330

Qy        100 LeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAsp 119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 CTCGGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGAC 390

Qy        120 TrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAsp 139
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        391 TGGATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGAC 450

Qy        140 IleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArg 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 ATTGACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGA 510

Qy        160 GlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArg 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GGCCCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGG 570

Qy        180 GluArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerTh 198
              ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        571 GAGCGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAAC 630

Qy        198 rProSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrTh 218
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        631 ACCCAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCAC 690

Qy        218 rProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              ||||||||||   ||||||||||||||||||||||||||||||||
Db        691 ACCAAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 736

RESULT 7
AAC98195
ID   AAC98195 standard; cDNA; 1655 BP.
XX
AC   AAC98195;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen nucleotide sequence SEQ ID NO:205.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
```

```
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; AAB53438.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 628-629; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1655 BP; 371 A; 415 C; 437 G; 425 T; 7 other;

Alignment Scores:
Pred. No.:                2.87e-83        Length:         1655
Score:                    958.50          Matches:        180
Percent Similarity:       84.19%          Conservative:   17
Best Local Similarity:    76.92%          Mismatches:     33
Query Match:              74.71%          Indels:         4
DB:                       21              Gaps:           2

US-09-972-970-4 (1-233) x AAC98195 (1-1655)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||   ||||||   ||||||||||
Db        277 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 330

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||||:::::||||||    ||||||      ||||||   ||||||||||
Db        331 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 390

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   |||||||||||||||||||||||||:::::::|||||||||||||||||
Db        391 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 450

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
                 |||||||||||||||||:::|||||||||||||||||||   :::||||||||||||
Db        451 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 510

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        511 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 570

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||||
Db        571 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 630

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::|||||||||   ||||||||||||||||||:::::|||||||||||||||||||
Db        631 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 690

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||   ||||||||||||   ||||||||||||   |||
Db        691 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 750

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                 :::|||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        751 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAG 810

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||||||||||||||||||   :::||||||||    ||||||||||
Db        811 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 870

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||||||   |||   |||         :::||||||         |||
Db        871 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 930

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||   ||||||||||||||||||||||||
Db        931 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 970

RESULT 8
ACC44092
ID   ACC44092 standard; DNA; 1567 BP.
XX
AC   ACC44092;
XX
DT   13-JUN-2003  (first entry)
XX
DE   Rat gamma-hydroxybutyrate receptor cDNA.
XX
KW   gene; rat; gamma-hydroxybutyrate; anti-epileptic; anxiolytic; ds;
KW   antineurodegeneration; antipsychotic; brain; dopamine; opioid; GABA;
KW   gamma-aminobutyric acid; diagnosis; epilepsy; anxiety; sleep disorder;
KW   behavioral disorder; neurodegeneration; Parkinson's disease; psychosis;
KW   schizophrenia.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             22..1560
FT                   /*tag=  a
FT                   /product= "gamma-hydroxybutyrate receptor"
XX
PN   WO200078948-A2.
XX
PD   28-DEC-2000.
XX
PF   19-JUN-2000; 2000WO-FR01687.
XX
PR   18-JUN-1999;   99FR-0007784.
XX
PA   (UYPA-) UNIV PASTEUR LOUIS.
XX
PI   Andriamanpandry C,  Maitre M;
XX
DR   WPI; 2001-091570/10.
DR   P-PSDB; ABP98695.
XX
PT   New rat brain gamma-hydroxybutyrate receptor and its encoding nucleic
PT   acid, useful for identifying agents for treating e.g. epilepsy -
XX
PS   Claim 5; Page 64-66; 66pp; French.
XX
CC   This sequence represents the cDNA sequence encoding a novel rat
```

```
CC   gamma-hydroxybutyrate receptor (GHBR) isolated from rat brain hypocampal
CC   cell cDNA library. The invention relates to the isolation of this novel
CC   sequence, fragments of it and homologues of the sequence except for those
CC   homologues that having GenBank accession numbers AAC 17120 (human
CC   tetraspan NET-4), AA615405, AA967250 or AI467230. GHBR is the receptor
CC   for gamma-hydroxybutyrate in the rat brain and is involved in regulating
CC   dopaminergic, opioid and GABA(gamma-aminobutyric acid)ergic activities.
CC   The nucleic acid that encodes the protein is used: (1) as primers or
CC   probes for detection/amplification, particularly for screening gene
CC   libraries to identify promoters and regulators of the GHBR gene; (2) for
CC   expression of recombinant polypeptides; and (3) to detect allelic
CC   variants, mutations, deletions, loss of heterozygosity or genetic
CC   abnormalities in the GHBR gene (for diagnosing diseases, or
CC   susceptibility, associated with abnormal expression of GHBR). The
CC   protein (or cells and transgenic animals expressing it) is used:
CC   (1) to screen for agents that interact with GHBR; (2) to study
CC   expression/activity of the receptor, including its interaction with
CC   other compounds; and (3) to raise antibodies (Ab) specific for GHBR.
CC   The Ab are used to detect/measure (I) in diagnostic immunoassays.
CC   The agents are used to prevent or treat diseases associated with abnormal
CC   expression or activity of GHBR, particularly those involving cerebral
CC   GABA(gamma-aminobutyric acid)ergic and/or dopaminergic activities, e.g.
CC   epilepsy, anxiety, sleep or behavioral disorders, withdrawal from
CC   addictive drugs, neurodegeneration (Parkinson's disease), psychoses,
CC   schizophrenia and regulation of secretion of hormones (growth hormone
CC   and prolactin) that are under dopaminergic control.
XX
SQ   Sequence 1567 BP; 281 A; 436 C; 480 G; 370 T; 0 other;

Alignment Scores:
Pred. No.:                5.22e-83        Length:          1567
Score:                    955.50          Matches:         193
Percent Similarity:       85.11%          Conservative:    7
Best Local Similarity:    82.13%          Mismatches:      29
Query Match:              74.47%          Indels:          8
DB:                       23              Gaps:            1

US-09-972-970-4 (1-233) x ACC44092 (1-1567)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||||||||||||||||||  ||||||:::||||||||||||||||||||||||||||||
Db        135 ATGCCCGGCAAGCATCAGCAATTCCAGGACCCTGAGGTTGGCTGCTGCGGGAAATACTTT 194

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTC 254

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAsp-LeuGlyGl 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 TGGGCCTGGGGTGAGAAGGGCGTTCTTTCCAACATCTCAGGGCGGACAGATCCAAGGCGG 314

Qy         60 yLeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAl 80
              |||||||||||||||||||||||||||||:::||||||:::||||||||||||||||||||
Db        315 TCTTGACCCCGTGTGGCTGTTTGTGGTGATTGGGGGAATCATGTCAGTGCTGGGCTTTGC 374

Qy         80 aGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLe 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 CGGCTGCATTGGGGCCCTCCGGGAAAACACCTTCCTGCTCAAATTTTTCTCTGTGTTCCT 434

Qy        100 uGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120
              |||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db        435 CGGCCTCATCTTCTTCCTGGAGCTGGCGGCCGGGATCCTGGCCTTCGTGTTCAAGGATTG 494

Qy        120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIl 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        495 GATCCGAGACCAACTTAACCTCTTCATCAACAACAATGTCAAAGCCTACCGGGACGATAT 554

Qy        140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGl 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 TGACCTTCAGAACCTTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGG 614

Qy        160 yProAsnAspTrpAsnLeuAsnIleTyr--PheAsnCysThrAspLeu-AsnProSerAr 179
              |||||||||||||||||||||||||     ||||||||||||||    ||||||||||||
Db        615 GCCCAATGACTGGAACCTCAACATCCGGACTTCAACTGCACTGACTTCAAACCCAAGCCG 674

Qy        179 gGluArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrPr 199
              ||||||||||||||||||||||||  |||  |||  |||||||||          :::
Db        675 CGAGCGCTGTGGGGTGCCCTTCTTCTGCTGGGTAAG-GACCCTGCGGAAGACGTCTCAAT 733

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
                 :::  :::          ||||||  |||||||||:::|||      ||||||
Db        734 ACCCATGTGGCTATACATCCGCTCAAACTG-AGCTGGAGCACAAGATTCATCTACACCAA 792

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArg 231
                              ||||||||||||   |||
Db        793 AGCTGGT------GGCCATTTGAGAAGTGGCTCAAGA 823

RESULT 9
AAC90015
ID   AAC90015 standard; cDNA; 1174 BP.
XX
AC   AAC90015;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Clone HE8EJ16 coding sequence #1.
XX
KW   Gene therapy; human; 4 transmembrane superfamily receptor protein;
KW   endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW   reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW   bacterial infection; viral; fungal; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070076-A1.
XX
PD   23-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13504.
XX
PR   19-MAY-1999;   99US-0135122.
PR   03-JUN-1999;   99US-0137797.
PR   11-JUN-1999;   99US-0138573.
PR   18-AUG-1999;   99US-0149447.
PR   28-JAN-2000;  2000US-0178770.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Fan P,  Roschke V,  Shi Y,  Komatsoulis GA;
PI   Rosen CA;
XX
DR   WPI; 2001-007502/01.
DR   P-PSDB; AAB49505.
XX
PT   Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT   superfamily receptor protein, useful for diagnosing, treating and/or
PT   preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS   Claim 1; Page 270; 297pp; English.
XX
CC   The present invention relates to isolated nucleic acids and proteins
CC   encoding human soluble 4 transmembrane superfamily receptor protein (see
CC   AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
CC   nucleic acid. The present sequence is useful for preventing, treating or
CC   ameliorating a medical condition and in diagnosing (susceptibility to) a
CC   pathological condition e.g. endocrine disorders e.g. Addison's disease,
CC   (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,
CC   cerebrovascular diseases, neural disorders e.g. Alzheimer's and
CC   Parkinson's disease, reproductive disorders, skin disorders e.g.
CC   psoriasis, renal system disorders e.g. nephritis, (auto)immune system
CC   disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.
CC   neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections
CC   caused by bacteria, viruses and fungi.
XX
SQ   Sequence 1174 BP; 306 A; 259 C; 291 G; 316 T; 2 other;
```

```
Alignment Scores:
Pred. No.:                1.39e-79        Length:          1174
Score:                    918.50          Matches:         179
Percent Similarity:       83.40%          Conservative:    17
Best Local Similarity:    76.17%          Mismatches:      34
Query Match:              71.59%          Indels:          6
DB:                       22              Gaps:            2

US-09-972-970-4 (1-233) x AAC90015 (1-1174)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||   ||||||   |||||||||
Db         93 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 146

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||:::::::||||||   ||||||      ||||||   |||||||||
Db        147 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 206

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   |||||||||||||||||||||||||:::::::|||||||||||||||
Db        207 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 266

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
                 ||||||||||||||||||:::|||||||||||||||||||   :::||||||||||||
Db        267 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 326

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        327 GGGTGCATTGGAGCGTA-CGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 385

Qy        101 GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120
              |||:::||||||| ||||||||||||      |||:::|||||||||||||||||||||||
Db        386 GGAATTATTTTCTTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTG 445

Qy        120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIl 140
              ||||:::|||||||||   ||||||||||||||||||:::::|||||||||||||||||||
Db        446 GATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACAT 505

Qy        140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGl 160
              |||||||||||||||||||||||   ||||||||||   ||||||||||   ||
Db        506 TGATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGG 565

Qy        160 yProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGl 180
              |   :::|||||||||||||||||||||||||||||||||||   |||   ||||||||
Db        566 AGCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGA 625

Qy        180 uArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrP 199
              |||||||||||||||||||||||||||   :::||||||||||    ||||||||||||
Db        626 GCGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACAC 685

Qy        199 roSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrP 219
              |||||||||||||||   |||   |||   :::||||||   |||
Db        686 TCAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACAC 745

Qy        219 roLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||   ||||||||||||||||||||||||||
Db        746 GAAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 786

RESULT 10
AAC90020
ID   AAC90020 standard; cDNA; 1178 BP.
XX
AC   AAC90020;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Clone HE8EL16 coding sequence #2.
XX
KW   Gene therapy; human; 4 transmembrane superfamily receptor protein;
KW   endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW   reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW   bacterial infection; viral; fungal; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070076-A1.
XX
PD   23-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13504.
XX
PR   19-MAY-1999;   99US-0135122.
PR   03-JUN-1999;   99US-0137797.
PR   11-JUN-1999;   99US-0138573.
PR   18-AUG-1999;   99US-0149447.
PR   28-JAN-2000; 2000US-0178770.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Fan P,  Roschke V,  Shi Y,  Komatsoulis GA;
PI   Rosen CA;
XX
DR   WPI; 2001-007502/01.
DR   P-PSDB; AAB49510.
XX
PT   Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT   superfamily receptor protein, useful for diagnosing, treating and/or
PT   preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS   Claim 1; Pages 274-275; 297pp; English.
XX
CC   The present invention relates to isolated nucleic acids and proteins
CC   encoding human soluble 4 transmembrane superfamily receptor protein (see
CC   AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
CC   nucleic acid. The present sequence is useful for preventing, treating or
CC   ameliorating a medical condition and in diagnosing (susceptibility to) a
CC   pathological condition e.g. endocrine disorders e.g. Addison's disease,
CC   (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,
CC   cerebrovascular diseases, neural disorders e.g. Alzheimer's and
CC   Parkinson's disease, reproductive disorders, skin disorders e.g.
CC   psoriasis, renal system disorders e.g. nephritis, (auto)immune system
CC   disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.
CC   neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections
CC   caused by bacteria, viruses and fungi.
XX
SQ   Sequence 1178 BP; 310 A; 261 C; 291 G; 316 T; 0 other;

Alignment Scores:
Pred. No.:                4.67e-77        Length:          1178
Score:                    892.50          Matches:         179
Percent Similarity:       82.70%          Conservative:    17
Best Local Similarity:    75.53%          Mismatches:      34
Query Match:              69.56%          Indels:          8
DB:                       22              Gaps:            2

US-09-972-970-4 (1-233) x AAC90020 (1-1178)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||   ||||||   |||||||||
Db         91 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 144

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||:::::::||||||   ||||||      ||||||   |||||||||
Db        145 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 204

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   |||||||||||||||||||||||||:::::::|||||||||||||||
Db        205 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 264

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
                 ||||||||||||||||||:::|||||||||||||||||||   :::||||||||||||
Db        265 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 324

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSer-ValPheLe 100
              |||||||||||||||   |||||||||||||||||||||||||||||||| ||||||||
Db        325 GGGTGCATTGGAGCGTA-CGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTCGTGTTCCT 383
```

```
Qy        100 u-GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAsp 119
              | |||:::|||||| ||||||||||||       |||:::||||||||||||||||||||
Db        384 CGGGAATTATTTTCTTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGAC 443

Qy        120 TrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAsp 139
              ||||||:::|||||||||   ||||||||||||||||||:::::|||||||||||||||
Db        444 TGGATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGAC 503

Qy        140 IleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArg 159
              |||||||||||||||||||||||||   ||||||||||||   ||||||||||||
Db        504 ATTGATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTT 563

Qy        160 GlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArg 179
              |||   :::|||||||||||||||||||||||||||||||||||||||   |||   ||||||
Db        564 GGAGCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGA 623

Qy        180 GluArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerTh 198
              |||||||||||||||||||||||||||||   :::|||||||||    ||||||||||
Db        624 GAGCGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAAC 683

Qy        198 rProSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrTh 218
              |   ||||||||||||||||   |||   |||         :::||||||         ||
Db        684 ACTCAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTAC 743

Qy        218 rProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |   |||||||||      ||||||||||||||||||||||||
Db        744 ACGAAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 786
```

RESULT 11
AAZ20854
ID AAZ20854 standard; cDNA; 1110 BP.
XX
AC AAZ20854;
XX
DT 09-DEC-1999 (first entry)
XX
DE Polynucleotide sequence of the dk329_1 clone.
XX
KW secreted protein; cDNA library; clone; transmembrane protein;
KW signal sequence cloning; hybridization cloning; gene therapy;
KW receptor; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   176..757
FT   /*tag= a
FT   /product= dk329_1
FT sig_peptide   326..424
FT   /*tag= b
FT mat_peptide   425..757
FT   /*tag= c
XX
PN WO9942470-A1.
XX
PD 26-AUG-1999.
XX
PF 18-FEB-1999; 99WO-US03458.
XX
PR 18-FEB-1998; 98US-0075038.
PR 17-FEB-1999; 99US-0251600.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Merberg D;
PI Treacy M, Agostino MJ, Steininger RJ;
XX
DR WPI; 1999-518580/43.
DR P-PSDB; AAY42381.
XX
PT New polynucleotides encoding human secreted proteins used for
PT therapeutic, diagnostic and research purposes.
XX
PS Claim 12; Page 101; 125pp; English.
XX
CC This is the polynucleotide sequence of the clone dk329_1, which was
CC isolated from a human fetal kidney cDNA library using methods which are
CC selective for cDNAs encoding secreted proteins, or by identification as
CC a secreted or transmembrane protein on the basis of computer analysis of
CC the amino acid sequence of the encoded protein (AAY42381).
CC The PNs and proteins of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC hematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumor invasion suppressor activity, and tumor
CC inhibition activity. The PNs are also stated to be useful for gene
CC therapy. Other activities include inhibiting the growth, infection or
CC function of bacteria, fungi, viruses and other parasites; effecting
CC bodily characteristics such as, e.g. weight, color, skin, etc.,
CC effecting biorhythms or circadian cycles; enhancing fertility; treatment
CC of depression; treatment of pain; hormonal or endocrine activity.
XX
SQ Sequence 1110 BP; 261 A; 273 C; 288 G; 285 T; 3 other;

Alignment Scores:

Pred. No.:	8.61e-69	Length:	1110
Score:	807.00	Matches:	163
Percent Similarity:	76.50%	Conservative:	16
Best Local Similarity:	69.66%	Mismatches:	30
Query Match:	62.90%	Indels:	26
DB:	20	Gaps:	3

US-09-972-970-4 (1-233) x AAZ20854 (1-1110)

```
Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::  |||||||||  ||||||   |||||||||
Db        176 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 229

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||:::::|||||   ||||||       ||||||   |||||||||
Db        230 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 289

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   |||||||||||||||||||||||||:::::|||||||||||||||||
Db        290 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 349

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||:::|||||||||||||||||||    :::||||||||||||
Db        350 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 409

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 469

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||       |||:::|||||||||||||||||||||||
Db        470 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 529

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::||||||||   ||||||||||||||||||:::::|||||||||||||||||||
Db        530 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 589

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              |||||||||||||||||||||||   |||||||||   |||
Db        590 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATAT-TCC------------------ 630

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                                                      |||   |||||||||
Db        631 ----------------------------------------------AATGCAAGTCGAGAG 645
```

```
Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||   :::|||||||||    ||||||||||||
Db        646 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 705

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              |||||||||||||||   |||  |||        :::||||||          |||
Db        706 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 765

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||     ||||||||||||||||||||||||
Db        766 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 805
```

RESULT 12
AAS59291
ID AAS59291 standard; cDNA; 1110 BP.
XX
AC AAS59291;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human cDNA encoding a secreted protein dk329_1.
XX
KW Human; secreted protein; ss; antiinflammatory; immunosuppressive;
KW nootropic; neuroprotective; antiarthritic; antimicrobial; vulnerary;
KW cytostatic; antidiabetic; virucide; antiinfertility; anticonvulsant;
KW vasotropic; antiparkinsonian; immunostimulant; dermatological;
KW antirheumatic; antitumor; antiulcer; osteopathic; tranquiliser;
KW cerebroprotective; cytokine; cell proliferation; cell differentiation;
KW immune deficiency; severe combined immunodeficiency; SCID; tumour;
KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;
KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; infection; cardiac disease;
KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;
KW food supplement.
XX
OS Homo sapiens.
XX
PN WO200175068-A2.
XX
PD 11-OCT-2001.
XX
PF 22-MAR-2001; 2001WO-US09369.
XX
PR 30-MAR-2000; 2000US-0539330.
PR 04-DEC-2000; 2000US-0729674.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;
PI Clark H, Fechtel K, Merberg D;
XX
DR WPI; 2001-639363/73.
DR P-PSDB; AAU39073.
XX
PT Secreted human proteins, useful as vaccine for treating various
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and
PT nervous system disorders (e.g. stroke) -
XX
PS Disclosure; Page 576-577; 619pp; English.
XX
CC The invention relates to novel human secreted proteins, the nucleic
CC acids encoding them. The protein may exhibit cytokine, cell proliferation
CC or cell differentiation activity or may induce production of other
CC cytokines in certain cell populations and may exhibit immune stimulating
CC or immune suppressing activity, which is useful for the treatment of
CC various immune deficiencies and disorders e.g. severe combined
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,
CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation. The proteins are also useful in the treatment of diseases
CC and disorders including tissue, skin and organ transplantation and in
CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,
CC in the treatment of burns, incisions and ulcers; as well as in treatment
CC of periodontal disease, osteoporosis or osteoarthritis, mediated by
CC inflammatory processes, diseases of the peripheral nervous system,
CC Alzheimer's, Parkinson's disease, Huntington's disease,
CC amylotrophic lateral sclerosis, and Shy-Drager syndrome, infections,
CC infarction of cardiac and central nervous system vessel e.g. stroke,
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The
CC protein, having activin- or inhibin-related activities is useful as a
CC contraceptive based on the ability of inhibins to decrease fertility in
CC female mammals and decrease spermatogenesis in male mammals. The
CC proteins and nucleic acids are also useful as food supplements. The
CC present sequence encodes a secreted protein of the invention.
XX
SQ Sequence 1110 BP; 261 A; 273 C; 288 G; 285 T; 3 other;

Alignment Scores:

Pred. No.:	8.61e-69	Length:	1110
Score:	807.00	Matches:	163
Percent Similarity:	76.50%	Conservative:	16
Best Local Similarity:	69.66%	Mismatches:	30
Query Match:	62.90%	Indels:	26
DB:	22	Gaps:	3

US-09-972-970-4 (1-233) x AAS59291 (1-1110)

```
Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||  ||||||     |||::::::  |||||||||  ||||||  ||||||||||
Db        176 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 229

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||||:::::|||||  ||||||      ||||||  ||||||||||
Db        230 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 289

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||  ||||||||||||||||||||||||||:::::|||||||||||||||||
Db        290 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 349

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||:::||||||||||||||||||||  :::||||||||||||
Db        350 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 409

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 469

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||
Db        470 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 529

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::|||||||||  ||||||||||||||||||:::::|||||||||||||||||||
Db        530 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 589

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||  |||||||||  |||
Db        590 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATAT-TCC----------------- 630

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                                                          |||  |||||||||
Db        631 ------------------------------------------AATGCAAGTCGAGAG 645

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||||||||||||||   :::|||||||||    ||||||||||||
Db        646 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 705

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              |||||||||||||||   |||  |||        :::||||||          |||
Db        706 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 765

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||     ||||||||||||||||||||||||
Db        766 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 805
```

```
RESULT 13
ABA90960
ID   ABA90960 standard; cDNA; 1110 BP.
XX
AC   ABA90960;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 169.
XX
KW   Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;
KW   immune disorder; bacterial infection; fungal infection; cancer; tumour;
KW   autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;
KW   osteoporosis; osteoarthritis; nervous system disorder; neuropathy;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;
KW   haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerary;
KW   ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;
KW   Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;
KW   neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.
XX
OS   Homo sapiens.
XX
PN   US2001039335-A1.
XX
PD   08-NOV-2001.
XX
PF   04-DEC-2000; 2000US-0729674.
XX
PR   10-APR-1997;   97US-126425P.
PR   04-DEC-1997;   97US-067454P.
PR   20-DEC-1997;   97US-068379P.
PR   02-JAN-1998;   98US-070346P.
PR   07-JAN-1998;   98US-070643P.
PR   08-JAN-1998;   98US-070755P.
PR   13-JAN-1998;   98US-071304P.
PR   22-JAN-1998;   98US-072134P.
PR   30-JAN-1998;   98US-073095P.
PR   18-FEB-1998;   98US-075038P.
PR   30-MAR-2000; 2000US-0539330.
PR   23-NOV-1998;   98US-0197886.
XX
PA   (JACO/) JACOBS K.
PA   (MCCO/) MCCOY J M.
PA   (LAVA/) LAVALLIE E R.
PA   (COLL/) COLLINS-RACIE L A.
PA   (EVAN/) EVANS C.
PA   (MERB/) MERBERG D.
PA   (TREA/) TREACY M.
PA   (AGOS/) AGOSTINO M J.
PA   (STEI/) STEININGER R J.
PA   (SPAU/) SPAULDING V.
PA   (WONG/) WONG G G.
PA   (CLAR/) CLARK H.
PA   (FECH/) FECHTEL K.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Agostino MJ,  Steininger RJ,  Spaulding V;
PI   Wong GG,  Clark H,  Fechtel K;
XX
DR   WPI; 2002-040725/05.
DR   P-PSDB; ABB55782.
XX
PT   New secreted proteins and encoding polynucleotides, useful in gene
PT   therapies, particularly for preventing or treating autoimmune
PT   disorders, cancer, graft-versus-host disease, wound, osteoporosis,
PT   stroke or inflammations  -
XX
PS   Disclosure; Page 307; 349pp; English.
XX
CC   The invention relates to isolated polynucleotides (ABA90876-ABA90968 and
CC   ABA90980) and encoded proteins (ABB55698-ABB55800), especially
CC   polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and
CC   proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in
CC   clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1
CC   are deposited with the American Type Culture Collection (ATCC) with
CC   accession number 98599. The polynucleotides and encoded polypeptides have
CC   cytostatic, anti-inflammatory, immunomodulator, vulnerary,
CC   neuroprotective, activin, inhibin, chemotactic, hAemostatic, thrombolytic
CC   and anti-inflammatory activity and acting as cytokine modulators,
CC   haematopoiesis regulators, tissue growth modulators and/or cadherin
CC   suppressors. The polypeptides and polynucleotides are useful in gene
CC   therapies, particularly for preventing, treating or ameliorating any of
CC   the following diseases: immune deficiency and disorders; e.g. bacterial
CC   or fungal infections, autoimmune disorders, cancer, systemic lupus
CC   erythematosus or graft-versus-host disease; myeloid or lymphoid cell
CC   deficiencies; wound, burns, incisions and ulcers, osteoporosis or
CC   osteoarthritis; central and peripheral nervous system diseases and
CC   neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's
CC   disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;
CC   haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis
CC   or systemic inflammatory response syndrome, ischaemia-reperfusion
CC   injury, endotoxin lethality, arthritis, inflammatory bowel disease or
CC   Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus
CC   foliaceus.
XX
SQ   Sequence 1110 BP; 261 A; 273 C; 288 G; 285 T; 3 other;

Alignment Scores:
Pred. No.:                8.61e-69        Length:          1110
Score:                    807.00          Matches:         163
Percent Similarity:       76.50%          Conservative:    16
Best Local Similarity:    69.66%          Mismatches:      30
Query Match:              62.90%          Indels:          26
DB:                       24              Gaps:            3

US-09-972-970-4 (1-233) x ABA90960 (1-1110)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||  ||||||     |||::::::  |||||||||  ||||||  ||||||||||
Db        176 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 229

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||||:::::|||||   ||||||      ||||||   |||||||||
Db        230 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 289

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||  ||||||||||||||||||||||||||||:::::|||||||||||||||
Db        290 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 349

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||:::||||||||||||||||||   :::||||||||||||||
Db        350 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 409

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 469

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||
Db        470 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 529

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::|||||||||   ||||||||||||||||:::::|||||||||||||||||||||
Db        530 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 589

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||   |||||||||   |||
Db        590 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATAT-TCC----------------- 630

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                                                              |||   |||||||||
Db        631 ----------------------------------------------AATGCAAGTCGAGAG 645

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||||||||||||||   :::||||||||    ||||||||||||
```

```
Db        646 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 705

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219

Db        706 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 765

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232

Db        766 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 805
```

RESULT 14
AAI72287
ID AAI72287 standard; cDNA; 864 BP.
XX
AC AAI72287;
XX
DT 15-APR-2002 (first entry)
XX
DE NET-4 antisense molecule.
XX
KW NET-4; antisense; modulator; neoplastic disease; cell proliferation;
KW gene expression; triple-helix; polymerase; transcription factor;
KW cancer therapy; hyperproliferation; tumour; growth; invasion;
KW metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200198350-A2.
XX
PD 27-DEC-2001.
XX
PF 05-JUN-2001; 2001WO-US18415.
XX
PR 07-JUN-2000; 2000US-209865P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Reinhard C, Jefferson AB, Winter JA, Randazzo F;
XX
DR WPI; 2002-147795/19.
XX
PT Novel NET-4 modulator useful for decreasing expression of NET-4 in a
PT mammalian cell and treating neoplastic disease, is selected from
PT antisense oligonucleotide, ribozyme, protein, polypeptide and a small
PT molecule -
XX
PS Claim 3; Page 50; 57pp; English.
XX
CC This sequence represents a NET-4 antisense molecule which acts as a
CC NET-4 modulator. The modulator is useful for decreasing the expression
CC of NET-4 in a mammalian cell, and for treating neoplastic disease,
CC such that the neoplastic disease is reduced in severity. Modulators
CC of NET-4 are also useful for regulating cell proliferation, and for
CC controlling gene expression through triple-helix formation which
CC promotes the ability of the double helix to open sufficiently for
CC the binding of polymerase, transcription factors or regulatory
CC molecules. A NET-4 modulator is useful as drug for supplementing
CC cancer therapeutics and other agents. It is also useful in
CC other diseases of hyperproliferation, and to inhibit tumour cell
CC growth, invasion or metastasis.
XX
SQ Sequence 864 BP; 162 A; 229 C; 249 G; 224 T; 0 other;

Alignment Scores:

Pred. No.:	2.76e-67	Length:	864
Score:	790.00	Matches:	139
Percent Similarity:	88.51%	Conservative:	15
Best Local Similarity:	79.89%	Mismatches:	18
Query Match:	61.57%	Indels:	2
DB:	24	Gaps:	1

US-09-972-970-4 (1-233) x AAI72287 (1-864)

```
Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20

Db        348 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 401

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40

Db        402 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 461

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60

Db        462 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 521

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80

Db        522 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 581

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100

Db        582 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 641

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120

Db        642 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 701

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140

Db        702 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 761

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160

Db        762 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 821

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174

Db        822 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGAT 863
```

RESULT 15
ABK35735
ID ABK35735 standard; cDNA; 1988 BP.
XX
AC ABK35735;
XX
DT 08-MAY-2002 (first entry)
XX
DE cDNA sequence #126 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; allergic condition; neurodegenerative disorder;
KW liver fibrosis; coagulation disorder; gene therapy; antimicrobial;
KW tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200177289-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10232.
XX
PR 06-APR-2000; 2000US-195605P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;
PI Clark HF, Fechtel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;
XX
DR WPI; 2002-179322/23.
XX
PT Six hundred and twenty three polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders

PT -
XX
PS Claim 1; Page 146; 393pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides
CC a method for producing proteins from these polynucleotide sequences.
CC The proteins are useful for identifying compounds that modulate their
CC activity and production. The sequences of the invention are
CC useful for treating diseases such as hyperproliferative disorders
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),
CC allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.
CC Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.
CC haemophilia), and tumours. The polynucleotide sequences of the
CC invention are also useful in gene therapy. ABK35610-ABK36232 represent
CC the cDNA sequences of the invention that encode for novel human
CC secreted proteins.
XX
SQ Sequence 1988 BP; 503 A; 434 C; 446 G; 604 T; 1 other;

Alignment Scores:

Pred. No.:	7.65e-64	Length:	1988
Score:	759.50	Matches:	146
Percent Similarity:	83.33%	Conservative:	14
Best Local Similarity:	76.04%	Mismatches:	30
Query Match:	59.20%	Indels:	2
DB:	24	Gaps:	1

US-09-972-970-4 (1-233) x ABK35735 (1-1988)

```
Qy         43 TrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAsp 62
              |||   |||||||||||||||||||||||||||||:::::|||||||||||||||   |||
Db          2 TGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGAC 61

Qy         63 ProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCys 82
              ||||||||||||||:::|||||||||||||||||   :::|||||||||||||||||||
Db         62 CCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGC 121

Qy         83 IleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeu 102
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        122 ATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTCTGTGTTCCTGGGAATT 181

Qy        103 IlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArg 122
              ||||||||||||||||||      |||:::|||||||||||||||||||||||||||:::
Db        182 ATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAA 241

Qy        123 AspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIleAspLeu 142
              ||||||||   |||||||||||||||||:::::|||||||||||||||||||||||||
Db        242 GACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTG 301

Qy        143 GlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGlyProAsn 162
              ||||||||||||||||   ||||||||||||   ||||||||||||   |||   :::
Db        302 CAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGAT 361

Qy        163 AspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGluArgCys 182
              ||||||||||||||||||||||||||||||||          :::            :::|||
Db        362 GATTGGAACCTAAATATTTACTTCAATTGTCCCGCAGAAGATGTCATCAACACTCAGTGT 421

Qy        183 GlyValProPheSerCysCysValArgAspProAla----MetSerSerThrProSerVa 201
              ||||||||||||||||||||   :::|||||||||    ||||||||||||   |||||
Db        422 GGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGT 481

Qy        201 lAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAl 221
              |||||||||   |||   |||          :::||||||          |||   |||||
Db        482 GGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGC 541

Qy        221 aValTrpAlaSerLeuArgSerGlyCysArgThr 232
              ||||      ||||||||||||||||||||||||
Db        542 TGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 575
```

Search completed: November 21, 2003, 15:34:34
Job time : 285 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 21, 2003, 13:58:33 ; Search time 3729 Seconds
(without alignments)
2556.166 Million cell updates/sec

Title: US-09-972-970-4
Perfect score: 1283
Sequence: 1 MPGKHQHFQEPEVGCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972970/runat_21112003_125335_27229/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972970_@CGN_1_1_3508_@runat_21112003_125335_27229 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1215.5	94.7	813	6	AX420468	AX420468 Sequence
	2	1215.5	94.7	2426	9	BC010405	BC010405 Homo sapi
	3	1215.5	94.7	3184	6	AX420466	AX420466 Sequence
	4	1174.5	91.5	1516	10	BC010346	BC010346 Mus muscu
	5	958.5	74.7	1405	9	AF065389	AF065389 Homo sapi
	6	958.5	74.7	1416	9	BC009704	BC009704 Homo sapi
	7	955.5	74.5	1567	6	AX061778	AX061778 Sequence
	8	954.5	74.4	1685	5	BC041304	BC041304 Xenopus l
	9	944.5	73.6	3175	10	AF121344	AF121344 Mus muscu
	10	925.5	72.1	1408	9	AF053455	AF053455 Homo sapi
	11	853.5	66.5	4445	9	AK024427	AK024427 Homo sapi
	12	807	62.9	1110	6	BD135990	BD135990 Secretory
	13	790	61.6	864	6	AX343015	AX343015 Sequence
	14	671	52.3	2428	10	BC025568	BC025568 Mus muscu
	15	671	52.3	2498	10	BC024611	BC024611 Mus muscu
	16	671	52.3	2500	10	BC026574	BC026574 Mus muscu
	17	654.5	51.0	1388	6	AX440923	AX440923 Sequence
	18	654.5	51.0	2553	9	AF311903	AF311903 Homo sapi
	19	651.5	50.8	2556	9	HSM801611	AL136638 Homo sapi
	20	579	45.1	1998	3	AK116798	AK116798 Ciona int
	21	570	44.4	2502	9	BC002920	BC002920 Homo sapi
	22	537	41.9	368	9	AF174603	AF174603 Homo sapi
c	23	501	39.0	171419	2	AC135142	AC135142 Rattus no
	24	501	39.0	234117	2	AC130985	AC130985 Rattus no
c	25	489	38.1	187478	2	AC123758	AC123758 Mus muscu
c	26	489	38.1	215066	2	AC136719	AC136719 Mus muscu
	27	476.5	37.1	852	6	AX247836	AX247836 Sequence
	28	476.5	37.1	2073	9	BC044244	BC044244 Homo sapi
	29	474.5	37.0	1995	10	BC024685	BC024685 Mus muscu
c	30	440.5	34.3	132832	9	AC091934	AC091934 Homo sapi
c	31	430	33.5	140468	2	AC017377	AC017377 Drosophil
c	32	430	33.5	174311	3	AC004758	AC004758 Drosophil
c	33	430	33.5	174832	3	AC092216	AC092216 Drosophil
	34	430	33.5	260673	3	AE003612	AE003612 Drosophil
c	35	426.5	33.2	168318	2	AC024042	AC024042 Homo sapi
	36	421.5	32.9	1816	3	BT004914	BT004914 Drosophil
	37	412.5	32.2	1216	3	AF274013	AF274013 Drosophil
c	38	404.5	31.5	22920	3	CET14G10	Z68880 Caenorhabdi
	39	402	31.3	102468	2	AC009984	AC009984 Drosophil
	40	402	31.3	106584	2	AC013977	AC013977 Drosophil
	41	402	31.3	140973	3	AC008140	AC008140 Drosophil
	42	402	31.3	159007	3	AC009219	AC009219 Drosophil
	43	402	31.3	225594	3	AE003688	AE003688 Drosophil
c	44	373.5	29.1	102298	10	AL671671	AL671671 Mouse DNA
	45	367	28.6	267	6	BD072489	BD072489 Secreted

ALIGNMENTS

RESULT 1

```
AX420468
LOCUS       AX420468                 813 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 3 from Patent WO0216603.
ACCESSION   AX420468
VERSION     AX420468.1  GI:21524616
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Leiby,K.R.
  TITLE     23228, a human tetraspanin family member and uses thereof
  JOURNAL   Patent: WO 0216603-A 3 28-FEB-2002;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      151 a    233 c    235 g    194 t
ORIGIN

Alignment Scores:
Pred. No.:                9.98e-96        Length:          813
Score:                    1215.50         Matches:         227
Percent Similarity:       96.60%          Conservative:    0
Best Local Similarity:    96.60%          Mismatches:      6
Query Match:              94.74%          Indels:          2
DB:                       6               Gaps:            1

US-09-972-970-4 (1-233) x AX420468 (1-813)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 60

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 120

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 180

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 240

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 300

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 360

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        361 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 420

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 480

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 540

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        541 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 600

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||||||      |||||
Db        601 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 660

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||   ||||||||||||||||||||||||||||||||
Db        661 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 703

RESULT 2
BC010405
LOCUS       BC010405                 2426 bp    mRNA    linear   PRI 10-JUN-2003
DEFINITION  Homo sapiens hypothetical protein MGC14859, mRNA (cDNA clone
            MGC:14859 IMAGE:3621871), complete cds.
ACCESSION   BC010405
VERSION     BC010405.1  GI:14714540
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2426)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2426)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 24 Row: k Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction, Similarity but not identity
            to protein.
FEATURES             Location/Qualifiers
     source          1. .2426
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:14859 IMAGE:3621871"
                     /tissue_type="Pancreas, adenocarcinoma"
                     /clone_lib="NIH_MGC_39"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .2426
                     /gene="MGC14859"
                     /note="synonym: FLJ00016"
                     /db_xref="LocusID:89852"
     CDS             61. .852
                     /codon_start=1
                     /product="MGC14859 protein"
                     /protein_id="AAH10405.1"
                     /db_xref="GI:14714541"
                     /db_xref="LocusID:89852"
                     /translation="MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG
                     EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL
                     IFFLELATGILAFVFKDWIRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARG
                     PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDVRLKLELEQQGFI
                     HTKGCVGQFEKWLQDNLIVVAGVFMGIALLQIFGICLAQNLEQME"
BASE COUNT      507 a    662 c    716 g    541 t
ORIGIN

Alignment Scores:
Pred. No.:              3.68e-95        Length:         2426
Score:                  1215.50         Matches:        227
Percent Similarity:     96.60%          Conservative:   0
Best Local Similarity:  96.60%          Mismatches:     6
Query Match:            94.74%          Indels:         2
DB:                     9               Gaps:           1

US-09-972-970-4 (1-233) x BC010405 (1-2426)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 120

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 180

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 240

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 300

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 360

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 420

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 480

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGC 540

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 600

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db        601 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 660

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 720

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||  ||||||||||||||||||||||||||||||||||
Db        721 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 763

RESULT 3
AX420466
LOCUS       AX420466                3184 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 1 from Patent WO0216603.
ACCESSION   AX420466
VERSION     AX420466.1  GI:21524614
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Leiby,K.R.
  TITLE     23228, a human tetraspanin family member and uses thereof
  JOURNAL   Patent: WO 0216603-A 1 28-FEB-2002;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3184
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             168. .980
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD35489.1"
                     /db_xref="GI:21524615"
                     /translation="MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG
                     EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL
                     IFFLELATGILAFVFKDWIRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARG
                     PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDVRLKLELEQQGFI
                     HTKGCVGQFEKWLQDNLIVVAGVFMGIALLQIFGICLAQNLVSDIKAVKANW"
BASE COUNT      675 a    883 c    945 g    681 t
ORIGIN

Alignment Scores:
Pred. No.:              5.08e-95        Length:         3184
Score:                  1215.50         Matches:        227
Percent Similarity:     96.60%          Conservative:   0
Best Local Similarity:  96.60%          Mismatches:     6
Query Match:            94.74%          Indels:         2
DB:                     6               Gaps:           1

US-09-972-970-4 (1-233) x AX420466 (1-3184)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 227

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 287

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 347

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 407

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 467
```

```
Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 527

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        528 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 587

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 647

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 707

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        708 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 767

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||      ||||||||||||||||||||||||||||||      |||||
Db        768 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 827

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              |||||||  ||||||||||||||||||||||||||||||
Db        828 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 870
```

RESULT 4
BC010346

```
LOCUS       BC010346               1516 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus RIKEN cDNA 2210021G21 gene, mRNA (cDNA clone MGC:6941
            IMAGE:2811935), complete cds.
ACCESSION   BC010346
VERSION     BC010346.1  GI:16307592
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1516)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 5 Row: f Column: 18.
FEATURES              Location/Qualifiers
     source           1. .1516
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="MGC:6941 IMAGE:2811935"
                      /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
                      old, gross tissue."
                      /clone_lib="NCI_CGAP_Mam5"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     gene             1. .1516
                      /gene="2210021G21Rik"
                      /db_xref="LocusID:74257"
                      /db_xref="MGI:1921507"
     CDS              136. .948
                      /codon_start=1
                      /product="RIKEN cDNA 2210021G21"
                      /protein_id="AAH10346.1"
                      /db_xref="GI:16307593"
                      /db_xref="LocusID:74257"
                      /translation="MPGKHQQFQDPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG
                      EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL
                      IFFLELAAGILAFVFKDWIRDQLNLFINNNVKAYRDDLDLQNLIDFAQEYWSCCGARG
                      PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDIRLKLELEQQGSI
                      YTKGCVGQFEKWLQDNLIVVAGVLVGIALLQIFGLCLAQNLVSDIKAVKANW"
BASE COUNT      277 a    417 c    461 g    361 t
ORIGIN

Alignment Scores:
Pred. No.:                7.12e-92        Length:          1516
Score:                    1174.50         Matches:         221
Percent Similarity:       95.32%          Conservative:    3
Best Local Similarity:    94.04%          Mismatches:      9
Query Match:              91.54%          Indels:          2
DB:                       10              Gaps:            1

US-09-972-970-4 (1-233) x BC010346 (1-1516)
```

```
Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||  ||||||:::||||||||||||||||||||||||||||||
Db        136 ATGCCCGGCAAGCACCAGCAATTCCAGGACCCTGAGGTCGGCTGCTGCGGGAAATACTTC 195

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATTGGCCTC 255

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCCAACATCTCGGCGCTGACAGATCTGGGCGGT 315

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CTTGACCCCGTGTGGCTGTTTGTGGTGGTGGGGGGAGTCATGTCAGTGTTGGGCTTCGCC 375
```

```
Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GGCTGCATTGGGGCCCTCCGGGAAAACACCTTCCTGCTCAAGTTCTTCTCTGTGTTCCTC 435

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GGCCTCATCTTCTTCCTGGAGCTGGCCGCGGGGATCCTGGCCTTCGTCTTCAAGGATTGG 495

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        496 ATCCGAGACCAACTTAATCTCTTCATCAACAACAATGTCAAAGCCTACCGGGATGATCTC 555

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 GACCTTCAGAACCTTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGTGCCCGAGGG 615

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCAAGCCGTGAG 675

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    |||||||||   ||
Db        676 CGCTGTGGGGTGCCCTTTTCCTGCTGTGTTAGGGACCCAGCGGAAGATGTCCTCAATACC 735

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              |||||||||||||   |||   ||||||||||||   ||||||||||||||   |||||
Db        736 CAGTGTGGCTATGACATCCGACTCAAACTGGAGCTCGAGCAGCAGGGCTCCATCTACACC 795

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
              ||||||||||||||||||||||||||||||   :::||||||
Db        796 AAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTACAAGACAACC 838

RESULT 5
AF065389
LOCUS       AF065389                1405 bp    mRNA    linear   PRI 28-APR-2000
DEFINITION  Homo sapiens tetraspan NET-4 mRNA, complete cds.
ACCESSION   AF065389
VERSION     AF065389.1  GI:3152702
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1405)
  AUTHORS   Serru,V., Dessen,P., Boucheix,C. and Rubinstein,E.
  TITLE     Sequence and expression of seven new tetraspans
  JOURNAL   Biochim. Biophys. Acta 1478 (1), 159-163 (2000)
  MEDLINE   20185353
   PUBMED   10719184
REFERENCE   2  (bases 1 to 1405)
  AUTHORS   Rubinstein,E., Serru,V. and Boucheix,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1998) INSERM U268, 14 av Paul Vaillant Couturier,
            Villejuif 94807, France
FEATURES             Location/Qualifiers
     source          1. .1405
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE ID 219547"
     CDS             348. .1154
                     /note="similar to Tspan-5; TM4SF"
                     /codon_start=1
                     /product="tetraspan NET-4"
                     /protein_id="AAC17120.1"
                     /db_xref="GI:3152703"
                     /translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
                     GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIF
                     FLELTAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGAD
                     DWNLNIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYT
                     KGCVPQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      309 a    357 c    387 g    352 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.6e-73       Length:          1405
Score:                    958.50        Matches:         180
Percent Similarity:       84.19%        Conservative:    17
Best Local Similarity:    76.92%        Mismatches:      33
Query Match:              74.71%        Indels:          4
DB:                       9             Gaps:            2

US-09-972-970-4 (1-233) x AF065389 (1-1405)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||  ||||||     |||::::::  |||||||||  ||||||  |||||||||
Db        348 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 401

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||::::::||||||  ||||||      ||||||  |||||||||
Db        402 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 461

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||  ||||||||||||||||||||||||::::::|||||||||||||||||
Db        462 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 521

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||:::||||||||||||||||||  :::|||||||||||||
Db        522 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 581

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 641

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||||
Db        642 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 701

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::||||||||||  ||||||||||||||||||:::::|||||||||||||||||||
Db        702 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 761

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||  ||||||||||||  ||||||||||||   |||
Db        762 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 821

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
              :::||||||||||||||||||||||||||||||||||||||||||  |||  |||||||||
Db        822 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAG 881

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              ||||||||||||||||||||||||||||||||||||||||||    :::||||||||||||
Db        882 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 941

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||||||   |||   |||          :::||||||          |||
Db        942 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 1001

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||      ||||||||||||||||||||||||
Db       1002 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 1041

RESULT 6
BC009704
LOCUS       BC009704                1416 bp    mRNA    linear   PRI 13-JAN-2003
DEFINITION  Homo sapiens, tetraspan 5, clone MGC:9300 IMAGE:3895933, mRNA,
            complete cds.
ACCESSION   BC009704
VERSION     BC009704.1  GI:16307230
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1416)
  AUTHORS   Strausberg,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 14 Row: h Column: 4
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21264582.
FEATURES             Location/Qualifiers
     source          1. .1416
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:9300 IMAGE:3895933"
                     /tissue_type="Pancreas, epithelioid carcinoma"
                     /clone_lib="NIH_MGC_70"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             334. .1140
                     /codon_start=1
                     /product="tetraspan 5"
                     /protein_id="AAH09704.1"
                     /db_xref="GI:16307231"
                     /db_xref="LocusID:10098"
                     /translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
                     GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIF
                     FLELTAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGAD
                     DWNLNIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYT
                     KGCVPQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      325 a    354 c    385 g    352 t
ORIGIN

Alignment Scores:
Pred. No.:              2.62e-73        Length:         1416
Score:                  958.50          Matches:        180
Percent Similarity:     84.19%          Conservative:   17
Best Local Similarity:  76.92%          Mismatches:     33
Query Match:            74.71%          Indels:         4
DB:                     9               Gaps:           2

US-09-972-970-4 (1-233) x BC009704 (1-1416)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||   ||||||   |||||||||
Db        334 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 387

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||:::::||||||   ||||||      ||||||   |||||||||
Db        388 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 447

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   |||||||||||||||||||||||||:::::|||||||||||||||||
Db        448 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 507

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
                 ||||||||||||||||||:::||||||||||||||||||   :::||||||||||||
Db        508 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 567

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        568 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 627

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||
Db        628 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 687

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::||||||||||   ||||||||||||||||||:::::|||||||||||||||||||
Db        688 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 747

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||||   ||||||||||||   ||||||||||||   |||
Db        748 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 807

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                 :::|||||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        808 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAG 867

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||||||||||||||||||||   :::|||||||||   ||||||||||||
Db        868 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 927

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||||||||   |||   |||          :::|||||||      |||
Db        928 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 987

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||      ||||||||||||||||||||||||
Db        988 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 1027

RESULT 7
AX061778
LOCUS       AX061778                1567 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 2 from Patent WO0078948.
ACCESSION   AX061778
VERSION     AX061778.1  GI:12539860
KEYWORDS    .
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Andriamanpandry,C. and Maitre,M.
  TITLE     Cloning, expression and characterisation of a cdna coding for a rat
            brain gamma-hydroxybutyrate (ghb) receptor
  JOURNAL   Patent: WO 0078948-A 2 28-DEC-2000;
            Universite Louis Pasteur de Strasbourg (FR)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
     CDS             22. .1560
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC25029.1"
                     /db_xref="GI:12539861"
                     /translation="MRAAAAGLGPGRLHAWAARRGLGRFPARVPRAAGGRSPCPASIS
                     NSRTLRLAAAGNTFCLASTLSSGCWEPCSWPSASGPGVRRAFFPTSQGGQIQGGLDPV
                     WLFVVIGGIMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELAAGILAFVFKDWIR
                     DQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIRTSTALTSNPSR
                     ERCGVPFFCWVRTLRKTSQYPCGYTSAQTELEHKIHLHQSWWPFEKWLKKPDGWPGLG
                     AIASFKMGIAGPNPRSTSRQERPTGKLPGVMATCLAHRAVVGASKDALPHSQWQGLWD
                     VCYDLSPRLCEHGTQEATEAGLGLNWGCTGAGLGLFTTELCVWGVHMCVCVCVCVCVR
                     VCLCLCVRVRGMHVCALVSTPGVSTPLLVRWWPFQSFKGDGARRVGHVPASPSLLWDV
                     SLCGLGACCLRPLHIHHDLEPAWSSPWPQCHSLEMGPRILSVSLSRLPLRTHTWQDGY
                     ERLN"
BASE COUNT      281 a    436 c    480 g    370 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:          5.36e-73        Length:         1567
Score:              955.50          Matches:        193
Percent Similarity: 85.11%          Conservative:   7
Best Local Similarity: 82.13%       Mismatches:     29
Query Match:        74.47%          Indels:         8
DB:                 6               Gaps:           1

US-09-972-970-4 (1-233) x AX061778 (1-1567)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||||||||||||||||||   ||||||:::|||||||||||||||||||||||||||||
Db        135 ATGCCCGGCAAGCATCAGCAATTCCAGGACCCTGAGGTTGGCTGCTGCGGGAAATACTTT 194

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTC 254

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAsp-LeuGlyGl 60
              ||||||||||||||||||||||||||||||||||||      ||||||    |||||
Db        255 TGGGCCTGGGGTGAGAAGGGCGTTCTTTCCAACATCTCAGGGCGGACAGATCCAAGGCGG 314

Qy         60 yLeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAl 80
              |||||||||||||||||||||||||:::|||||||:::||||||||||||||||||||||
Db        315 TCTTGACCCCGTGTGGCTGTTTGTGGTGATTGGGGGAATCATGTCAGTGCTGGGCTTTGC 374

Qy         80 aGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLe 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 CGGCTGCATTGGGGCCCTCCGGGAAAACACCTTCCTGCTCAAATTTTTCTCTGTGTTCCT 434

Qy        100 uGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120
              ||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db        435 CGGCCTCATCTTCTTCCTGGAGCTGGCGGCCGGGATCCTGGCCTTCGTGTTCAAGGATTG 494

Qy        120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIl 140
              |||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        495 GATCCGAGACCAACTTAACCTCTTCATCAACAACAATGTCAAAGCCTACCGGGACGATAT 554

Qy        140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGl 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 TGACCTTCAGAACCTTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGG 614

Qy        160 yProAsnAspTrpAsnLeuAsnIleTyr--PheAsnCysThrAspLeu-AsnProSerAr 179
              |||||||||||||||||||||||||     |||||||||||||||    ||||||||||||
Db        615 GCCCAATGACTGGAACCTCAACATCCGGACTTCAACTGCACTGACTTCAAACCCAAGCCG 674

Qy        179 gGluArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrPr 199
              |||||||||||||||||||||||||   |||   |||   |||||||||          :::
Db        675 CGAGCGCTGTGGGGTGCCCTTCTTCTGCTGGGTAAG-GACCCTGCGGAAGACGTCTCAAT 733

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
                 :::   :::         ||||||   |||||||||:::|||      ||||||
Db        734 ACCCATGTGGCTATACATCCGCTCAAACTG-AGCTGGAGCACAAGATTCATCTACACCAA 792

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArg 231
                              ||||||||||||   |||
Db        793 AGCTGGT------GGCCATTTGAGAAGTGGCTCAAGA 823

RESULT 8
BC041304
LOCUS       BC041304                1685 bp    mRNA    linear   VRT 14-JAN-2003
DEFINITION  Xenopus laevis, Similar to transmembrane 4 superfamily member 9,
            clone IMAGE:4683897, mRNA, partial cds.
ACCESSION   BC041304
VERSION     BC041304.1  GI:27735442
KEYWORDS    .
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1685)
  AUTHORS   Klein,S. and Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-2002) National Institutes of Health, Xenopus Gene
            Collection (XGC), National Institute of Child Health and Human
            Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
            20892-7510, USA
  REMARK    NIH-MGC Project
COMMENT     Contact: XGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Igor Dawid
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 94 Row: g Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES              Location/Qualifiers
     source           1. .1685
                      /organism="Xenopus laevis"
                      /mol_type="mRNA"
                      /db_xref="taxon:8355"
                      /clone="IMAGE:4683897"
                      /tissue_type="Embryo, stage 31/32, Xenopus"
                      /clone_lib="NICHD_XGC_Emb4"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              <1. .926
                      /codon_start=3
                      /product="Similar to transmembrane 4 superfamily member 9"
                      /protein_id="AAH41304.1"
                      /db_xref="GI:27735443"
                      /translation="FGSTRSVGLIYCFHLAPSRAVSPDPPRCQPGNLPSCSGSMSGKH
                      YKGPEVSCCIKYFIFGFNVIFWLLGLTFLGVGLWAWSEKGVLSNISSITDLGGFDPVW
                      LFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDWIKD
                      QLQFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRER
                      CGVPFSCCTKDPAEDVINTQCGYDVRQKPELDQQETIHTKGCVPQFEKWLQDNLTIVA
                      GVFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      456 a    332 c    383 g    514 t
ORIGIN

Alignment Scores:
Pred. No.:          7.13e-73        Length:         1685
Score:              954.50          Matches:        175
Percent Similarity: 84.26%          Conservative:   23
Best Local Similarity: 74.47%       Mismatches:     33
Query Match:        74.40%          Indels:         4
DB:                 5               Gaps:           2

US-09-972-970-4 (1-233) x BC041304 (1-1685)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::  |||||||||   ||||||   |||||||||
Db        120 ATGTCTGGCAAG------CACTACAAGGGACCCGAGGTCAGCTGCTGCATCAAGTACTTC 173

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||||:::::|||||||:::||||||       ||||||   :::||||||
Db        174 ATATTCGGCTTCAATGTCATCTTCTGGCTCTTGGGACTTACGTTTCTTGGAGTTGGTCTG 233

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   ||||||||||||||||||||||||||||||:::::|||||||||||||||
Db        234 TGGGCATGGAGTGAAAAGGGTGTGCTTTCCAACATTTCCTCTATCACAGATCTTGGGGGT 293

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
                 |||||||||||||||||:::|||||||||||||||||||   :::||||||||||||
Db        294 TTTGATCCAGTGTGGCTGTTTCTGGTAGTCGGAGGGGTCATGTTTATCCTGGGCTTTGCA 353
```

```
Qy          81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         354 GGCTGTATTGGGGCACTCCGTGAGAACACCTTCCTGTTGAAATTTTTTTCTGTGTTTCTT 413

Qy         101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
               |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||
Db         414 GGAATTATTTTCTTTCTGGAGCTCACGGCGGGAGTTCTTGCCTTCGTGTTCAAAGACTGG 473

Qy         121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
               |||:::|||||||||   |||||||||||||||||:::::|||||||||||||||||||
Db         474 ATTAAAGACCAACTGCAGTTCTTTATTAACAACAACATCAGAGCCTACAGAGATGACATC 533

Qy         141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
               ||||||||||||||||||||||||   ||||||||||||   ||||||||||||   |||
Db         534 GACTTACAGAACCTTATTGACTTCACACAGGAATATTGGCAGTGCTGCGGGGCGTTCGGA 593

Qy         161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
               :::|||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         594 GCTGATGATTGGAATTTAAACATTTATTTCAACTGCACCGACTCTAACGCGAGCCGAGAG 653

Qy         181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
               ||||||||||||||||||||||||||   :::|||||||||   |||       |||
Db         654 AGATGTGGGGTGCCATTTTCCTGCTGCACAAAAGACCCTGCTGAGGATGTAATAAACACA 713

Qy         199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
               |||||||||||||||||||       :::   :::|||   :::|||       :::|||
Db         714 CAGTGTGGCTATGATGTCAGACAAAAGCCTGAACTGGATCAACAGGAGACAATTCACACA 773

Qy         219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
               |||||||||      |||||||||:::|||||||||   |||
Db         774 AAAGGCTGTGTCCCACAGTTTGAGAAATGGCTGCAGGATAACC 816

RESULT 9
AF121344
LOCUS       AF121344                3175 bp    mRNA    linear   ROD 01-FEB-2000
DEFINITION  Mus musculus tetraspanin Tspan-5 (Tspan5) mRNA, complete cds.
ACCESSION   AF121344
VERSION     AF121344.1  GI:6841032
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3175)
  AUTHORS   Garcia-Frigola,C., de Lecea,L. and Soriano,E.
  TITLE     Mouse Tspan-5 cDNA cloning
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3175)
  AUTHORS   Garcia-Frigola,C., de Lecea,L. and Soriano,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1999) Dept. of Animal and Plant Cell Biology,
            University of Barcelona, Av. Diagonal 645, Barcelona 08028, Spain
FEATURES             Location/Qualifiers
     source          1. .3175
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Swiss Webster/NIH"
                     /db_xref="taxon:10090"
     gene            1. .3175
                     /gene="Tspan5"
     CDS             388. .1194
                     /gene="Tspan5"
                     /note="member of transmembrane 4 superfamily"
                     /codon_start=1
                     /product="tetraspanin Tspan-5"
                     /protein_id="AAF28869.1"
                     /db_xref="GI:6841033"
                     /translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
                     GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIF
                     FLELTAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGAD
                     DWNLNIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYT
                     KGCVPQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      667 a    811 c    840 g    857 t
```

```
ORIGIN

Alignment Scores:
Pred. No.:                1.1e-71        Length:          3175
Score:                    944.50         Matches:         178
Percent Similarity:       83.76%         Conservative:    18
Best Local Similarity:    76.07%         Mismatches:      34
Query Match:              73.62%         Indels:          4
DB:                       10             Gaps:            2

US-09-972-970-4 (1-233) x AF121344 (1-3175)

Qy           1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
               |||   ||||||      |||::::::   |||||||||   ||||||   |||||||||
Db         388 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 441

Qy          21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
               :::|||||||||||:::::|||||   ||||||      ||||||   |||||||||
Db         442 ATTTTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACGTTTCTTGGAATCGGACTG 501

Qy          41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
               |||||||||   |||||||||||||||||||||||||||||:::::|||||||||||||||
Db         502 TGGGCGTGGAATGAAAAAGGTGTCCTCTCCAACATCTCGTCCATCACCGACCTCGGTGGC 561

Qy          61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
               ||||||||||||||||||:::|||||||||||||||||||   :::|||||||||||||
Db         562 TTTGACCCAGTGTGGCTTTTCCTTGTGGTGGGAGGAGTGATGTTCATTCTGGGGTTTGCA 621

Qy          81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         622 GGGTGCATCGGAGCACTTCGGGAAAACACCTTTCTTCTCAAGTTTTTTTCTGTGTTCCTG 681

Qy         101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
               |||:::|||||||||||||||||||      |||:::|||||||||||||||||||||||
Db         682 GGGATTATTTTCTTCCTGGAACTCACTGCTGGGGTGTTGGCATTTGTTTTCAAAGACTGG 741

Qy         121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
               |||:::|||||||||   |||||||||||||||||:::::|||||||||||||||||||
Db         742 ATCAAAGACCAGCTGTATTTCTTTATTAACAACAACATCAGAGCCTACAGAGATGACATT 801

Qy         141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
               ||||||||||||||||||||||||   ||||||||||||   ||||||||||||   |||
Db         802 GATCTACAGAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 861

Qy         161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
               :::|||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         862 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGCCGAGAG 921

Qy         181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
               ||||||||||||||||||||||||||   :::|||||||||   |||       |||
Db         922 CGATGCGGTGTGCCATTTTCCTGCTGCACTAAAGACCCCGCGGAAGATGTCATCAACACT 981

Qy         199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
               |||||||||||||||   |||   |||         ::::::|||         |||
Db         982 CAGTGTGGCTATGATGCCAGGCAGAAACCAGAAGTTGACCAACAGATTGTAATCTACACA 1041

Qy         219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
               |||||||||      ||||||||||||||   ||||||
Db        1042 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGCTACAGGACA 1081

RESULT 10
AF053455
LOCUS       AF053455                1408 bp    DNA     linear   PRI 03-NOV-1998
DEFINITION  Homo sapiens tetraspan TM4SF (TSPAN-5) gene, complete cds.
ACCESSION   AF053455
VERSION     AF053455.1  GI:2995864
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1408)
  AUTHORS   Todd,S.C., Doctor,V.S. and Levy,S.
```

```
  TITLE     Sequences and expression of six new members of the
            tetraspanin/TM4SF family
  JOURNAL   Biochim. Biophys. Acta 1399 (1), 101-104 (1998)
  MEDLINE   98390278
   PUBMED   9714763
REFERENCE   2  (bases 1 to 1408)
  AUTHORS   Todd,S.C., Doctor,V.S. and Levy,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1998) Medicine, Stanford, 300 Pasteur Dr,
            Stanford, CA 94305, USA
FEATURES             Location/Qualifiers
     source          1. .1408
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            352. .1146
                     /gene="TSPAN-5"
     CDS             352. .1146
                     /gene="TSPAN-5"
                     /note="transmembrane 4 superfamily"
                     /codon_start=1
                     /product="tetraspan TM4SF"
                     /protein_id="AAC69712.1"
                     /db_xref="GI:2995865"
                     /translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
                     GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLSVFLGIIFFLEL
                     TAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNL
                     NIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYTKGCV
                     PQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      309 a    354 c    384 g    345 t     16 others
ORIGIN

Alignment Scores:
Pred. No.:              1.81e-70        Length:         1408
Score:                  925.50          Matches:        176
Percent Similarity:     82.48%          Conservative:   17
Best Local Similarity:  75.21%          Mismatches:     33
Query Match:            72.14%          Indels:         8
DB:                     9               Gaps:           3

US-09-972-970-4 (1-233) x AF053455 (1-1408)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              |||   ||||||      |||::::::   |||||||||   ||||||   |||||||||
Db        352 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 405

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              :::|||||||||||:::::||||||   ||||||       ||||||   ||||||||||
Db        406 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 465

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              |||||||||   |||||||||||||||||||||||||:::::|||||||||||||||||
Db        466 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 525

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
                |||||||||||||||||||:::|||||||||||||||||||   :::||||||||||||
Db        526 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 585

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTT-----------TCTGTGTTCCTG 633

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              |||:::|||||||||||||||||||       |||:::|||||||||||||||||||||||
Db        634 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 693

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              |||:::|||||||||   |||||||||||||||||||:::::|||||||||||||||||||
Db        694 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 753

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
              ||||||||||||||||||||||   ||||||||||   ||||||||||   |||
Db        754 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 813

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                :::|||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        814 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAG 873

Qy        181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
              |||||||||||||||||||||||||||||||||||||||   :::||||||||||    ||||||||||||
Db        874 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 933

Qy        199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
              ||||||||||||||||   |||   |||          :::||||||          |||
Db        934 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 993

Qy        219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
              |||||||||       ||||||||||||||||||||||||
Db        994 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 1033

RESULT 11
AK024427
LOCUS       AK024427               4445 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens mRNA for FLJ00016 protein, partial cds.
ACCESSION   AK024427
VERSION     AK024427.1  GI:10440362
KEYWORDS    fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4445)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
  TITLE     The nucleotide sequence of a long cDNA clone isolated from human
            spleen
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 4445)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/NEDO, Tel:81-438-52-3913,
            Fax:81-438-52-3914)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Kazusa DNA Research Institute.
FEATURES             Location/Qualifiers
     source          1. .4445
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="as00016"
                     /tissue_type="spleen"
                     /dev_stage="adult"
                     /note="vector:pBluescriptII SK plus"
     gene            1. .4445
                     /gene="FLJ00016"
     CDS             <2041. .3090
                     /gene="FLJ00016"
                     /note="For this clone, GeneMark analysis triggered an
                     alert for spurious CDS split. The result of GeneMark
                     analysis of the nucleotide sequence of this clone is
                     accessible through http://www.kazusa.or.jp/NEDO.
                     Start codon is not identified."
                     /codon_start=1
                     /product="FLJ00016 protein"
                     /protein_id="BAB15717.1"
                     /db_xref="GI:10440363"
                     /translation="GVLGGGDGVWDDASRHGGNGVARLWGWQDSPPVPALKRGGLCRT
                     SGASVGGLLQTWGPAVSAATGLGPAPASQVVWRKPQSLNQYPCPTPSLFLASLGSGKG
                     ELSVPSLAPLCLHQPGGWAQACGGGEVATHLAVPALSQTALYSCSLGLFPNRVRLLGW
                     AWLTWGSPLPPQWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVL
                     NTQCGYDVRLKLELEQQGFIHTKGCVGQFEKWLQDNLIVVAGVFMGIALLQIFGICLA
```

```
                    QNLVSDIKAVKANWSKWNDDFENHWLTPTISEVLSTAGPQQNSLTGAPGPAPPSRHVF
                    FGLGGLYPEPTFKNW"
BASE COUNT      890 a   1278 c   1321 g    956 t
ORIGIN

Alignment Scores:
Pred. No.:              1.13e-63        Length:         4445
Score:                  853.50          Matches:        227
Percent Similarity:     25.33%          Conservative:   0
Best Local Similarity:  25.33%          Mismatches:     6
Query Match:            66.52%          Indels:         663
DB:                     9               Gaps:           2

US-09-972-970-4 (1-233) x AK024427 (1-4445)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 179

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 239

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 299

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 359

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 419

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 479

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        480 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 539

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyr---------------------- 152
              |||||||||||||||||||||||||||||||
Db        540 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACGTGAGTCCAGTGTCCAGCCTGGGA 599

Qy        152 ------------------------------------------------------------ 152

Db        600 CACCTGTAGGAGGCGCCTTCCTGGCAGACGAATGAGGGAGCAAGTTCCCTGTGATGGGAC 659

Qy        152 ------------------------------------------------------------ 152

Db        660 CATCTCCTTACCCACCTGGGCTAGCGGGCCCCAGGAGAGAGACCCAGGAGTGTCCTCGTC 719

Qy        152 ------------------------------------------------------------ 152

Db        720 CGAGGAGAGCCAGCCCCTGCTGATCCTTATGGGTCCCCATTTGCCTAGTGCCATAGGGCC 779

Qy        152 ------------------------------------------------------------ 152

Db        780 ACGGCCTTGGGAAACAGAGGGCACGGGCTCCACACTAGCAGCTCCCAGGCCAGGGCCTAC 839

Qy        152 ------------------------------------------------------------ 152

Db        840 GACACACGGGAGAATGGAGGAAATGTGACATAAAATATAGATTTCTACTTCTGCTAAACC 899

Qy        152 ------------------------------------------------------------ 152

Db        900 TTGGGCCCACGGTATTGGCTGCAGCTATGTTGCTGCTGCCCCTTCACTGGTACAGAGGCT 959

Qy        152 ------------------------------------------------------------ 152

Db        960 CTCTGTTTGCCATTGTTCCCAACACTCTGTATTTTATTATATCTGTCAGCCCTTATCTCT 1019

Qy        152 ------------------------------------------------------------ 152

Db       1020 CATTTATACTATGTTTGGCCTCTGTGTATATTAGAGTTTAGACCCCTGATGGAAGGTCTG 1079

Qy        152 ------------------------------------------------------------ 152

Db       1080 CCCTCAAACTTGCAAACTTAACCCCAGTGAAGAGCAGCCAGAGAAGAGTTCCTGAAACCC 1139

Qy        152 ------------------------------------------------------------ 152

Db       1140 TATCTGCCTCCCACCTGCACCATCTCGCCCACTCACCCCTAACCACAAACACACAGTCAC 1199

Qy        152 ------------------------------------------------------------ 152

Db       1200 ACACTGTCACCCACAGTCACACCCACACCTAGTCACACACTCGCAGGCACACTTGATCAC 1259

Qy        152 ------------------------------------------------------------ 152

Db       1260 ACCTACACCTACACATGGTCCCACAGTCACACATGAGTCACACCCGATCACACCCACACA 1319

Qy        152 ------------------------------------------------------------ 152

Db       1320 CGGCGACAGGCCCTTGGCACCATTCACCACCAAGAGTCAGCTCTGGATCCAGAGGTGAAG 1379

Qy        152 ------------------------------------------------------------ 152

Db       1380 CAGAGCTGGTGCCTGCTTCAAGCAGCTCTCAACCCAGGGCAGAGGCAGACATAGCCAGGC 1439

Qy        152 ------------------------------------------------------------ 152

Db       1440 AGTTATGATGCAGTTCATTCATTCATTCAGCACAGGATTGTTAAGGGCCTACCATATGCC 1499

Qy        152 ------------------------------------------------------------ 152

Db       1500 GGACCCAGAGCCAGGCCCTGAGGATTCGGCGGGGAAGAAAGCTGGGACCTGTCCTCTCAG 1559

Qy        152 ------------------------------------------------------------ 152

Db       1560 GGCTGGCAGTCGGGAAGAGACAGATGGTGAACAAGCAGAACAAACAAATGTGTGCATGTA 1619

Qy        152 ------------------------------------------------------------ 152

Db       1620 CATACGTGTGTATGCAAAATTGTACATTGTGGTAAGAGTCAAAATGAAGGGTGGAGAATG 1679

Qy        152 ------------------------------------------------------------ 152

Db       1680 GGAAGGCCTCTGCGTGGAGAGCTTGAAGCCTCCCAAAGGAGGAGAAGGAAGAGAAGGGAG 1739

Qy        152 ------------------------------------------------------------ 152

Db       1740 AGTGTTCTGTAGCAGGAGCAGCTGGCGCTGAGCAGGTGCCAGGAGGCAGGTGGAGGCTGA 1799

Qy        152 ------------------------------------------------------------ 152

Db       1800 GCCCTGCCCTGGGACCAGCACATGCAGGGCATGTGGCAGGCCTGACTCCTCCAGCTCTCT 1859

Qy        152 ------------------------------------------------------------ 152

Db       1860 GGCTCGGCCACGGGGCCTGCCTGTGTCTGCTGCCGAGCTTGAGCTCCCCCAGCCCTGTGG 1919

Qy        152 ------------------------------------------------------------ 152

Db       1920 GCTGCTCTCCAGCCCCTGCGCTGCTGCTCAGTGAGGGCTGTGTGCACCCGTGTGCCCCAC 1979

Qy        152 ------------------------------------------------------------ 152

Db       1980 AAGACCATGGTTCATGGCAGATCCCAAACTCGCCATTGCTCCTGCCTAGGGGACAGCTTA 2039

Qy        152 ------------------------------------------------------------ 152

Db       2040 GGGTGTTCTGGGTGGGGGTGATGGGGTGTGGGATGATGCCAGCCGGCATGGAGGAAATGG 2099
```

```
Qy      152 ------------------------------------------------------------ 152

Db     2100 GGTGGCACGGCTATGGGGGTGGCAGGATTCCCCTCCTGTCCCTGCACTGAAGAGAGGAGG 2159

Qy      152 ------------------------------------------------------------ 152

Db     2160 GCTTTGTAGAACCTCTGGGGCAAGTGTGGGAGGCCTGCTGCAGACATGGGGCCCAGCGGT 2219

Qy      152 ------------------------------------------------------------ 152

Db     2220 CTCTGCTGCCACAGGGCTTGGCCCAGCGCCTGCCTCTCAGGTAGTCTGGAGGAAGCCACA 2279

Qy      152 ------------------------------------------------------------ 152

Db     2280 GTCATTGAACCAGTATCCTTGTCCCACTCCCTCCCTTTTTCTAGCCTCTCTTGGGTCGGG 2339

Qy      152 ------------------------------------------------------------ 152

Db     2340 GAAGGGGGAGCTCAGTGTCCCCTCCCTTGCACCCCTCTGCCTCCACCAGCCTGGAGGTTG 2399

Qy      152 ------------------------------------------------------------ 152

Db     2400 GGCCCAGGCCTGTGGGGGTGGGGAGGTGGCCACCCACTTGGCTGTCCCAGCCCTGTCCCA 2459

Qy      152 ------------------------------------------------------------ 152

Db     2460 GACAGCCCTGTATTCCTGCAGCCTGGGCTTGTTCCCAAACAGGGTGAGGCTCCTGGGGTG 2519

Qy      153 -------------------------------------TrpSerCysCysGlyAlaArgGl 160
                                                 ||||||||||||||||||||||||
Db     2520 GGCCTGGCTCACCTGGGGCTCTCCCCTGCCCCCACAGTGGTCTTGCTGCGGAGCCCGAGG 2579

Qy      160 yProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGl 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2580 CCCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGA 2639

Qy      180 uArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrP 199
            ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db     2640 GCGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACAC 2699

Qy      199 roSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrP 219
            ||||||||||       |||||||||||||||||||||||||||||||||      ||||
Db     2700 CCAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACAC 2759

Qy      219 roLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
            ||||||||   |||||||||||||||||||||||||||||||||
Db     2760 CAAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 2803
```

```
RESULT 12
BD135990
LOCUS       BD135990             1110 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Secretory proteins and polynucleotides encoding the same.
ACCESSION   BD135990
VERSION     BD135990.1  GI:23230935
KEYWORDS    JP 2002504488-A/2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1110)
  AUTHORS   Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A.C., Merberg,D.,
            Treacy,M., Agostino,M.J. and Ii,R.J.S.
  TITLE     Secretory proteins and polynucleotides encoding the same
  JOURNAL   Patent: JP 2002504488-A 2 12-FEB-2002;
            GENETICS INSTITUTE INC
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002504488-A/2
            PD   12-FEB-2002
            PF   18-FEB-1999 JP 2000532422
            PR   18-FEB-1998 US     60/075038,17-FEB-1999 US     09/251600 PI
            KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A COLLINS   PI
            RACIE,
            PI   DAVID MERBERG,MAURICE TREACY,MICHAEL J AGOSTINO,ROBERT J   PI
            STEININGER II
            PC
            C07K14/47,A61K38/00,A61P43/00,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12N15/09,C12P21/02,C12Q1/68,A61K37/02,C12N5/00,C12N15/00 CC
            Secretory proteins and polynucleotides encoding the same FH   Key
                     Location/Qualifiers
            FT   source          1. .1110
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .1110
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      261 a    273 c    288 g    285 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              2.18e-60       Length:          1110
Score:                  807.00         Matches:         163
Percent Similarity:     76.50%         Conservative:    16
Best Local Similarity:  69.66%         Mismatches:      30
Query Match:            62.90%         Indels:          26
DB:                     6              Gaps:            3

US-09-972-970-4 (1-233) x BD135990 (1-1110)

Qy        1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
            |||   ||||||      |||::::::  |||||||||   ||||||   ||||||||||
Db      176 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 229

Qy       21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
            :::|||||||||||||:::::::||||||    ||||||      ||||||   |||||||
Db      230 ATATTTGGCTTCAATGTCATATTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 289

Qy       41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
            |||||||||   ||||||||||||||||||||||||||||||::::::||||||||||||||
Db      290 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 349

Qy       61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
               ||||||||||||||||||:::|||||||||||||||||||    :::|||||||||||
Db      350 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 409

Qy       81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      410 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 469

Qy      101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
            |||:::||||||||||||||||||  ||||||:::||||||||||||||||||||||||||
Db      470 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 529

Qy      121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
            |||:::|||||||||   ||||||||||||||||||:::::::|||||||||||||||||||
Db      530 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 589

Qy      141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
            ||||||||||||||||||||||||    |||||||||   |||
Db      590 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATAT-TCC------------------ 630

Qy      161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
                                                           |||   |||||||||
Db      631 --------------------------------------------AATGCAAGTCGAGAG 645

Qy      181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
            |||||||||||||||||||||||||||||      :::|||||||||   |||||||:|||
Db      646 CGATGTGGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACT 705

Qy      199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
             ||||||||||||||   |||   ||||      :::||||||         |||
Db      706 CAGTGTGGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACG 765

Qy      219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
```

```
Db        766 AAAGGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 805

RESULT 13
AX343015
LOCUS       AX343015                 864 bp    DNA     linear   PAT 12-JAN-2002
DEFINITION  Sequence 1 from Patent WO0198350.
ACCESSION   AX343015
VERSION     AX343015.1  GI:18152215
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Reinhard,C., Jefferson,A.B., Winter,J.A. and Randazzo,F.
  TITLE     Compositions and methods for treating neoplastic disease using net
            -4 modulators
  JOURNAL   Patent: WO 0198350-A 1 27-DEC-2001;
            CHIRON CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .864
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="NET-4 oligonucleotide used in cell proliferation
                     assay on SW620 cells"
BASE COUNT      162 a    229 c    249 g    224 t
ORIGIN

Alignment Scores:
Pred. No.:                4.71e-59        Length:         864
Score:                    790.00          Matches:        139
Percent Similarity:       88.51%          Conservative:   15
Best Local Similarity:    79.89%          Mismatches:     18
Query Match:              61.57%          Indels:         2
DB:                       6               Gaps:           1

US-09-972-970-4 (1-233) x AX343015 (1-864)

Qy          1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
Db        348 ATGTCCGGGAAG------CACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTC 401

Qy         21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db        402 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTG 461

Qy         41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db        462 TGGGCATGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGC 521

Qy         61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80
Db        522 TTTGACCCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCA 581

Qy         81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
Db        582 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTG 641

Qy        101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
Db        642 GGAATTATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGG 701

Qy        121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
Db        702 ATCAAAGACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATT 761

Qy        141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
Db        762 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGA 821

Qy        161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174
```

```
Db        822 GCTGATGATTGGAACCTAAATATTTACTTCAATTGCACAGAT 863

RESULT 14
BC025568
LOCUS       BC025568                 2428 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus DNA segment, Chr 14, ERATO Doi 226, expressed, mRNA
            (cDNA clone MGC:36595 IMAGE:5322531), complete cds.
ACCESSION   BC025568
VERSION     BC025568.1  GI:19343883
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2428)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2428)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 56 Row: j Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction.
FEATURES             Location/Qualifiers
```

```
     source          1. .2428
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:36595 IMAGE:5322531"
                     /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
                     ductal carcinoma. 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .2428
                     /gene="D14Ertd226e"
                     /note="synonym: MGC36554"
                     /db_xref="LocusID:52588"
                     /db_xref="MGI:1196325"
     CDS             64. .876
                     /codon_start=1
                     /product="tetraspanin similar to TM4SF9"
                     /protein_id="AAH25568.1"
                     /db_xref="GI:19343884"
                     /db_xref="LocusID:52588"
                     /translation="MHYYRYSNAEVSCWYKYLLFSYNIVFWLAGVVFLGVGLWAWSEK
                     GVLSDLTKVTRLHGIDPVVLVLMVGVVMFTLGFAGCVGALRENICLLKFFCGAIVLIF
                     FLELAVAVLAFLFQDWVRDRFREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPE
                     DWDLNVYFNCSGASYSREKCGVPFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDEFIFT
                     KGCIQALEGWLPRNIYIVAGVFIAISLLQIFGIFLARTLISDIEAVKAGHHF"
BASE COUNT      517 a    640 c    624 g    647 t
ORIGIN

Alignment Scores:
Pred. No.:              2.86e-48        Length:         2428
Score:                  671.00          Matches:        134
Percent Similarity:     66.93%          Conservative:   34
Best Local Similarity:  53.39%          Mismatches:     60
Query Match:            52.30%          Indels:         23
DB:                     10              Gaps:           5

US-09-972-970-4 (1-233) x BC025568 (1-2428)

Qy          2 ProGlyLysHisGlnHis---------PheGlnGluProGluValGlyCysCysGlyLys 18
              |||   :::      |||         :::         ||||||  |||     |||
Db         52 CCTTCTCAGAAGATGCACTATTATAGATACTCGAACGCCGAGGTCAGCTGCTGGTACAAG 111

Qy         19 TyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIle 38
              |||   ||||||  :::|||||||||||||||:::   |||   :::||||||   :::
Db        112 TACCTGCTCTTCAGCTACAATATCGTCTTTTGGCTGGCTGGAGTTGTCTTCCTTGGAGTC 171

Qy         39 GlyLeuTrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeu 58
              ||||||||||||||   ||||||||||||||||||:::::::::   :::|||   |||
Db        172 GGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCGACCTCACCAAGGTGACCCGGTTG 231

Qy         59 GlyGlyLeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGly 78
              |||:::|||||||||   |||   ::::::||||||   ||||||      ||||||
Db        232 CATGGAATCGACCCCGTCGTGCTGGTCTTGATGGTGGGCGTGGTGATGTTCACACTGGGA 291

Qy         79 PheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerVal 98
              ||||||||||:::|||||||||||||||||      ||||||||||||||
Db        292 TTCGCAGGCTGTGTCGGGGCCCTCCGAGAGAACATCTGCCTGCTCAAGTTTTTCTGTGGG 351

Qy         99 PheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLys 118
              :::   ||||||||||||||||||||||||      :::|||||||||:::|||:::
Db        352 GCCATTGTGCTCATCTTCTTCCTGGAACTGGCCGTGGCCGTGTTGGCCTTTTTATTCCAA 411

Qy        119 AspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAsp 138
              ||||||:::||||||:::          |||      :::|||:::|||:::|||||||||
Db        412 GACTGGGTGAGAGACCGGTTCCGGGAATTCTTCGAGAGCAACATCAAGTCCTATCGGGAT 471

Qy        139 AspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAla 158
              ||||||||||||||||||||||||||      |||:::          ||||||||||||
Db        472 GACATCGACCTGCAGAACCTCATTGACTCCCTTCAGAAAGCTAATCAGTGCTGCGGGGCT 531

Qy        159 ArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSer 178
              ||||||  ||||||:::||||||:::||||||||||||:::      :::   |||
Db        532 TACGGCCCTGAAGACTGGGACCTCAATGTCTACTTCAACTGCAGTGGTGCCAGCTACAGC 591

Qy        179 ArgGluArgCysGlyValProPheSerCysCysValArgAspProAlaMet-SerSer-- 197
              ||||||:::|||||||||||||||||||||||||||||   |||||||||      |||
Db        592 CGAGAGAAATGTGGGGTACCCTTCTCCTGCTGTGTGCCAGATCCTGCACAAAAAGTCGTG 651

Qy        198 -ThrProSerValAlaMetMetSerGly------------------------------- 206
              |||   ||||||||||||||||||||||
Db        652 AACACACAGTGTGGCTATGATGTCCGGATTCAGCTGAAGAGCAAGTGGGATGAGTTCATC 711

Qy        207 ------------SerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaVa 222
                          ::::::      |||   :::      |||   |||         ::
Db        712 TTTACAAAAGGATGCATCCAGGCTCTGGAAGGCTGGCTGCCCAGGAACA---TCTACATT 768

Qy        222 lTrp------AlaSerLeuArgSerGlyCys 230
              :|||      :::||||||   |||   |||
Db        769 GTGGCTGGTGTCTTCATTGCCATCTCACTGC 799

RESULT 15
BC024611
LOCUS       BC024611               2498 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus DNA segment, Chr 14, ERATO Doi 226, expressed, mRNA
            (cDNA clone MGC:28503 IMAGE:4188261), complete cds.
ACCESSION   BC024611
VERSION     BC024611.1  GI:19354053
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2498)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2498)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
```

```
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 37 Row: e Column: 12
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..2498
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:28503 IMAGE:4188261"
                     /tissue_type="Salivary gland, 10 week old female mouse"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..2498
                     /gene="D14Ertd226e"
                     /note="synonym: MGC36554"
                     /db_xref="LocusID:52588"
                     /db_xref="MGI:1196325"
     CDS             55..867
                     /codon_start=1
                     /product="tetraspanin similar to TM4SF9"
                     /protein_id="AAH24611.1"
                     /db_xref="GI:19354054"
                     /db_xref="LocusID:52588"
                     /translation="MHYYRYSNAEVSCWYKYLLFSYNIVFWLAGVVFLGVGLWAWSEK
                     GVLSDLTKVTRLHGIDPVVLVLMVGVVMFTLGFAGCVGALRENICLLKFFCGAIVLIF
                     FLELAVAVLAFLFQDWVRDRFREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPE
                     DWDLNVYFNCSGASYSREKCGVPFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDEFIFT
                     KGCIQALEGWLPRNIYIVAGVFIAISLLQIFGIFLARTLISDIEAVKAGHHF"
BASE COUNT      542 a    644 c    631 g    681 t
ORIGIN

Alignment Scores:
Pred. No.:              2.96e-48       Length:          2498
Score:                  671.00         Matches:         134
Percent Similarity:     66.93%         Conservative:    34
Best Local Similarity:  53.39%         Mismatches:      60
Query Match:            52.30%         Indels:          23
DB:                     10             Gaps:            5

US-09-972-970-4 (1-233) x BC024611 (1-2498)

Qy          2 ProGlyLysHisGlnHis---------PheGlnGluProGluValGlyCysCysGlyLys 18
              |||  :::     |||        :::       ||||||  |||        |||
Db         43 CCTTCTCAGAAGATGCACTATTATAGATACTCGAACGCCGAGGTCAGCTGCTGGTACAAG 102

Qy         19 TyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIle 38
              |||   ||||||   :::|||||||||||||:::   |||   :::||||||   :::
Db        103 TACCTGCTCTTCAGCTACAATATCGTCTTTTGGCTGGCTGGAGTTGTCTTCCTTGGAGTC 162

Qy         39 GlyLeuTrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeu 58
              |||||||||||||||   |||||||||||||||||:::::::::   :::|||   |||
Db        163 GGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCGACCTCACCAAGGTGACCCGGTTG 222

Qy         59 GlyGlyLeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGly 78
                 |||:::|||||||||   |||   ::::::||||||   ||||||      ||||||
Db        223 CATGGAATCGACCCCGTCGTGCTGGTCTTGATGGTGGGCGTGGTGATGTTCACACTGGGA 282

Qy         79 PheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerVal 98
              ||||||||||||:::||||||||||||||||||         |||||||||||||||
Db        283 TTCGCAGGCTGTGTCGGGGCCCTCCGAGAGAACATCTGCCTGCTCAAGTTTTTCTGTGGG 342

Qy         99 PheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLys 118
                 :::   |||||||||||||||||||||||         :::|||||||||:::|||:::
Db        343 GCCATTGTGCTCATCTTCTTCCTGGAACTGGCCGTGGCCGTGTTGGCCTTTTTATTCCAA 402

Qy        119 AspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAsp 138
              ||||||:::||||||:::           |||       :::|||:::|||:::||||||
Db        403 GACTGGGTGAGAGACCGGTTCCGGGAATTCTTCGAGAGCAACATCAAGTCCTATCGGGAT 462

Qy        139 AspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAla 158
              |||||||||||||||||||||||||||      |||:::         ||||||||||||
Db        463 GACATCGACCTGCAGAACCTCATTGACTCCCTTCAGAAAGCTAATCAGTGCTGCGGGGCT 522

Qy        159 ArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSer 178
                 ||||||   ||||||:::||||||:::||||||||||||:::      :::   |||
Db        523 TACGGCCCTGAAGACTGGGACCTCAATGTCTACTTCAACTGCAGTGGTGCCAGCTACAGC 582

Qy        179 ArgGluArgCysGlyValProPheSerCysCysValArgAspProAlaMet-SerSer-- 197
              ||||||:::|||||||||||||||||||||||||||   |||||||||         |||
Db        583 CGAGAGAAATGTGGGGTACCCTTCTCCTGCTGTGTGCCAGATCCTGCACAAAAAGTCGTG 642

Qy        198 -ThrProSerValAlaMetMetSerGly------------------------------- 206
                |||   ||||||||||||||||||
Db        643 AACACACAGTGTGGCTATGATGTCCGGATTCAGCTGAAGAGCAAGTGGGATGAGTTCATC 702

Qy        207 -------------SerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaVa 222
                           ::::::   |||   :::      |||   |||        ::
Db        703 TTTACAAAAGGATGCATCCAGGCTCTGGAAGGCTGGCTGCCCAGGAACA---TCTACATT 759

Qy        222 lTrp------AlaSerLeuArgSerGlyCys 230
              :|||      :::||||||   |||   |||
Db        760 GTGGCTGGTGTCTTCATTGCCATCTCACTGC 790

Search completed: November 21, 2003, 16:37:00
Job time : 3739 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 16:37:04 ; Search time 5185 Seconds
(without alignments)
11896.772 Million cell updates/sec

Title: US-09-972-970-2
Perfect score: 2538
Sequence: 1 ccacgcgtccggccgcagcc.....aaaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

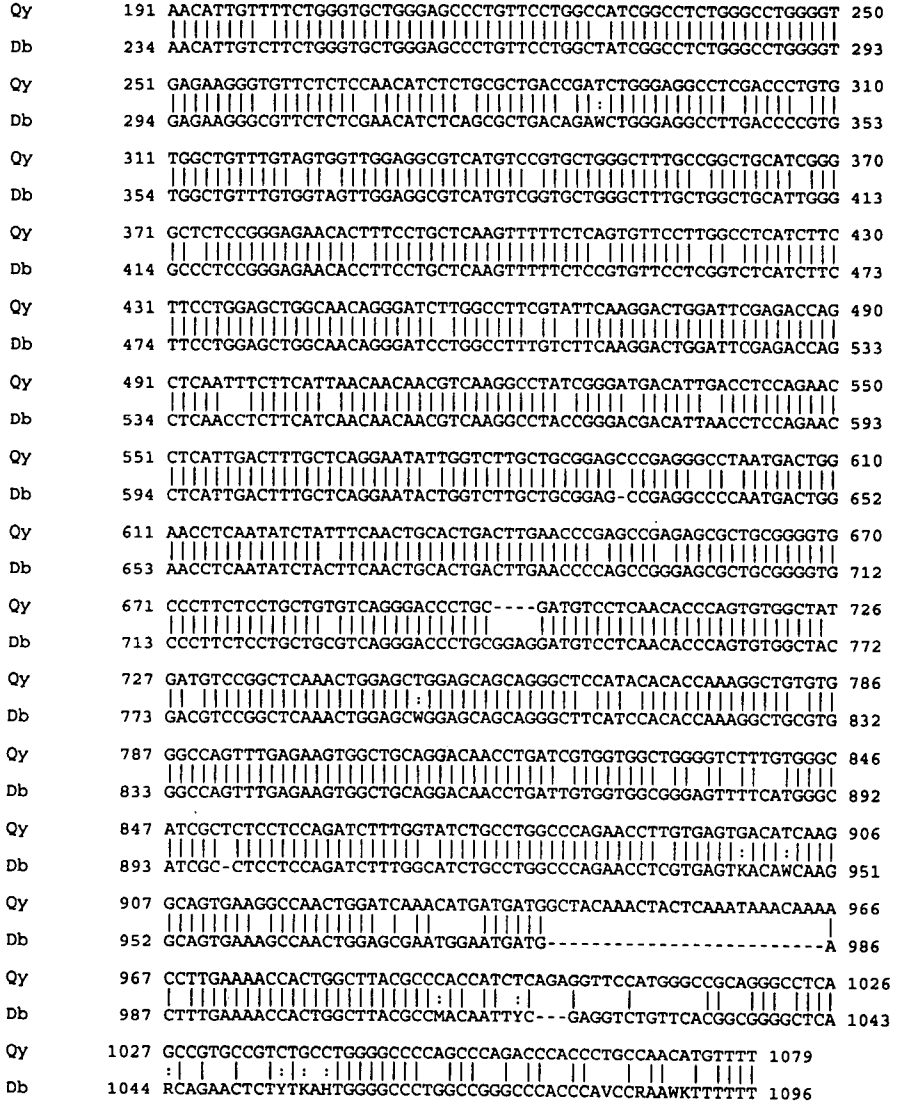
SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	796	31.4	1201	9	AL528502	AL528502 AL528502
2	771.8	30.4	1194	9	AL529630	AL529630 AL529630
3	743.8	29.3	1494	11	AK008761	AK008761 Mus muscu
4	692.2	27.3	957	9	AL543914	AL543914 AL543914
5	677.8	26.7	1005	14	BY708665	BY708665 BY708665
6	663.8	26.2	938	13	BU856801	BU856801 AGENCOURT
7	655.2	25.8	849	10	BE615772	BE615772 601279927
8	632.2	24.9	1032	12	BM563474	BM563474 AGENCOURT
9	622	24.5	1201	9	AL556793	AL556793 AL556793
10	621	24.5	1028	10	BG477727	BG477727 602521422
11	619.2	24.4	901	14	CA454987	CA454987 AGENCOURT
12	611.6	24.1	888	13	BX447619	BX447619 BX447619
13	606.4	23.9	781	12	BG770931	BG770931 602719148
14	602.4	23.7	718	12	BI914325	BI914325 603182305
15	568.4	22.4	902	10	BG478644	BG478644 602525532
16	559.2	22.0	970	12	BI909709	BI909709 603070671
17	552.2	21.8	782	12	BI158921	BI158921 602921828
18	543.4	21.4	1020	12	BI557863	BI557863 603236858
19	541.2	21.3	654	12	BI829529	BI829529 603080360
20	533.2	21.0	912	12	BI915506	BI915506 603177449
21	528.2	20.8	832	10	BF981395	BF981395 602309182
22	521	20.5	594	10	BF043938	BF043938 BP250022B
23	519.8	20.5	657	12	BI850548	BI850548 imageqc_1
24	515.8	20.3	916	13	BQ924634	BQ924634 AGENCOURT
25	503.6	19.8	588	10	BE615323	BE615323 601280719
26	489.6	19.3	538	10	BE683087	BE683087 181563 MA
27	489.6	19.3	960	10	BF119347	BF119347 601757362
28	480.6	18.9	1108	13	BQ072716	BQ072716 AGENCOURT
29	478.6	18.9	685	14	BY734818	BY734818 BY734818
30	472.6	18.6	769	10	BF122628	BF122628 601760347
31	468.2	18.4	600	12	BI987393	BI987393 3201-06 M
32	463	18.2	634	14	CB215495	CB215495 NISC_np06
33	451.8	17.8	704	10	BG422301	BG422301 602448952
34	449.4	17.7	3141	11	AK013350	AK013350 Mus muscu
35	443.4	17.5	714	10	BG017161	BG017161 de61f11.y
36	443.4	17.5	883	14	CA986012	CA986012 AGENCOURT
37	442	17.4	576	14	CB607991	CB607991 AMGNNUC:N
38	440	17.3	932	14	CB321353	CB321353 AGENCOURT
39	435.4	17.2	920	13	BQ941931	BQ941931 AGENCOURT
40	431.6	17.0	553	14	CB613921	CB613921 AMGNNUC:N
41	430.2	17.0	933	12	BM475708	BM475708 AGENCOURT
42	430	16.9	721	10	BF983242	BF983242 602305881
43	425.6	16.8	753	14	CA315923	CA315923 UI-M-FW0-
44	418.4	16.5	2108	11	AK033543	AK033543 Mus muscu
45	416.8	16.4	1116	11	AK015705	AK015705 Mus muscu

ALIGNMENTS

RESULT 1
AL528502
LOCUS AL528502 1201 bp mRNA linear EST 23-MAY-2003
DEFINITION AL528502 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC029YJ10 5-PRIME, mRNA sequence.
ACCESSION AL528502
VERSION AL528502.2 GI:31066352
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12791995.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3528.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC029DE05QP1&cluster=3528.f. Contact : Feng Liang Email : fliang@lifetech.com URL :

```
          http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CS0DC029DE05QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC029YJ10"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      236 a    331 c    340 g    262 t     32 others
ORIGIN

  Query Match           31.4%;  Score 796;  DB 9;  Length 1201;
  Best Local Similarity   91.3%;  Pred. No. 9e-97;
  Matches  860;  Conservative    6;  Mismatches   71;  Indels    5;  Gaps    2;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||||||||||||| || | || |
Db         64 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 123

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||||| ||| ||||||||||||| ||||||
Db        124 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 183

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||| |||||||||||||||||||||||||||||||||||||||
Db        184 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 243

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              |||||||| |||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        244 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 303

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
              |||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| |||
Db        304 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTG 363

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
              |||||||||| || |||||||||||||||||||| |||||||||||||| |||||||| |||
Db        364 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 423

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
              || |||||||||||||| |||||||||||||||||||| |||||||| || |||||||||
Db        424 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 483

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
              |||||||||||||||||||||| ||||||| || ||||||||||||||||||||||||
Db        484 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 543

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
              |||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||||
Db        544 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 603

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
              |||||||||||||||||||||| ||||||||||||||:|||||| || |||||||||
Db        604 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGASCCCGAGGCCCCAATGACTGG 663

Qy        611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
              |||||||||||||| |||||||||||||||||||||| ||||| ||||||||||||||||
Db        664 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 723

Qy        671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
              |||||||||||||| ||||||||||||||    |||||||||||||||||||||||||||
Db        724 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTAC 783

Qy        727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTG 786
              || |||||||||||||||||||||||||||||||||| ||| ||||||||||||||| |||
Db        784 GACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTG 843

Qy        787 GGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGC 846
              |||||||||||||||||||||||||||||||||||||||||| |||||||| || |||||  |||||
Db        844 GGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGC 903

Qy        847 ATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAG 906
              |||||  :|||||||||||||||| |||||||||||||||||||| ||||||:|||||||
Db        904 ATCGCCTYCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTKACATCAAA 963

Qy        907 GCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAA 948
              |:|||||| |||||||||| ::||   ||||||| ||   ||
Db        964 GSAGTGAAAGCCAACTGGAGSMAATGGGATGATGACTTTGAA 1005


RESULT 2
AL529630
LOCUS       AL529630               1194 bp    mRNA    linear   EST 23-MAY-2003
DEFINITION  AL529630 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
            cDNA clone CS0DD005YB20 5-PRIME, mRNA sequence.
ACCESSION   AL529630
VERSION     AL529630.2  GI:31067473
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12793123.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            3528.f For more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DD005DA10QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DD005DA10QP1.
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DD005YB20"
                     /tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      237 a    331 c    336 g    253 t     37 others
ORIGIN

  Query Match           30.4%;  Score 771.8;  DB 9;  Length 1194;
  Best Local Similarity   86.0%;  Pred. No. 1.5e-93;
  Matches  923;  Conservative   11;  Mismatches  104;  Indels   35;  Gaps    6;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              ||||||||:||| ||||||||||||||||||||||||||||||||||| || | || |
Db         54 GCCGCAGCSGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 113

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||||| ||| ||||||||||||| ||||||
Db        114 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 173

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              ||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        174 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 233
```



```
RESULT 3
AK008761
LOCUS       AK008761               1494 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male stomach cDNA, RIKEN full-length enriched
            library, clone:2210021G21 product:F-BOX PROTEIN FBX23 (FRAGMENT)
            homolog [Homo sapiens], full insert sequence.
ACCESSION   AK008761
VERSION     AK008761.1  GI:12843154
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1494)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
```

```
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES                Location/Qualifiers
     source             1. .1494
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="FANTOM_DB:2210021G21"
                        /db_xref="MGI:1907404"
                        /db_xref="taxon:10090"
                        /clone="2210021G21"
                        /sex="male"
                        /tissue_type="stomach"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="adult"
     CDS                136. .948
                        /note="unnamed protein product; F-BOX PROTEIN FBX23
                        (FRAGMENT) homolog [Homo sapiens] (SPTR|Q9UKB9, evidence:
                        FASTY, 86.9%ID, 99.1%length, match=363)
                        putative"
                        /codon_start=1
                        /protein_id="BAB25880.1"
                        /db_xref="GI:12843155"
                        /db_xref="MGI:1921507"
                        /translation="MPGKHQQFQDPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG
                        EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL
                        IFFLELAAGILAFVFKDWIRDQLNLFINNNVKAYRDDLDLQNLIDFAQEYWSCCGARG
                        PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDIRLKLELEQEGSI
                        YTKGCVGQFEKWLQDNLIVVAGVLVGIALLQIFGLCLAQNLVSDIKAVKANW"
     polyA_signal       1479. .1484
                        /note="putative"
     polyA_site         1494
                        /note="putative"
BASE COUNT      257 a    414 c    463 g    360 t
ORIGIN

  Query Match             29.3%;  Score 743.8;  DB 11;  Length 1494;
  Best Local Similarity   89.0%;  Pred. No. 6.9e-90;
  Matches  816;  Conservative    0;  Mismatches   97;  Indels    4;  Gaps    1;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              ||||||||| |||||||||||||| |||| |||| || || |||||||||| ||||||  |
Db         29 GCCGCAGCCACCGGGCTAGGCCCGGGGCTGCTCCAGGCCTGGGCGGCCCGCAGAGGGCTG 88

Qy         72 ATCCCGGCCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGC 131
                || | | ||||||| |||||||||| |||| || ||| ||||||||||| |||||||
Db         89 GGCCGGTCACCGGCTCGGGTTCCCGGGACGGCAGGTGGCCGCTCACCATGCCCGGCAAGC 148

Qy        132 ACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 191
              ||||||| |||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        149 ACCAGCAATTCCAGGACCCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 208

Qy        192 ACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTG 251
              ||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        209 ACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATTGGCCTCTGGGCCTGGGGTG 268

Qy        252 AGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGT 311
              ||||||| |||||||||||||||||| |||||||| |||||||| || || ||||| ||||
Db        269 AGAAGGGCGTTCTCTCCAACATCTCGGCGCTGACAGATCTGGGCGGTCTTGACCCCGTGT 328

Qy        312 GGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGG 371
              |||||||||| ||||| || || |||||||| ||| ||||||| |||||||||||| ||||
Db        329 GGCTGTTTGTGGTGGTGGGGGGAGTCATGTCAGTGTTGGGCTTCGCCGGCTGCATTGGGG 388

Qy        372 CTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCT 431
              | |||||||| ||||| |||||||||||||| ||||| |||||||| ||||||||||||
Db        389 CCCTCCGGGAAAACACCTTCCTGCTCAAGTTCTTCTCTGTGTTCCTCGGCCTCATCTTCT 448

Qy        432 TCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGC 491
              |||||||||||||  | |||||| |||||||||| |||||||| ||||| |||||||| |
Db        449 TCCTGGAGCTGGCCGCGGGGATCCTGGCCTTCGTCTTCAAGGATTGGATCCGAGACCAAC 508

Qy        492 TCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACC 551
              | ||| ||||||| |||||||| ||||| ||||| |||||||| | ||||| |||||||
Db        509 TTAATCTCTTCATCAACAACAATGTCAAAGCCTACCGGGATGATCTCGACCTTCAGAACC 568

Qy        552 TCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGA 611
              | || |||||||||||||||| |||||||||| || |||||||||| ||||||||||
Db        569 TTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGTGCCCGAGGGCCCAATGACTGGA 628

Qy        612 ACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGC 671
              |||||||||||||| |||||||||||||||||||||||| ||||| |||||||| |||||||
Db        629 ACCTCAATATCTACTTCAACTGCACTGACTTGAACCCAAGCCGTGAGCGCTGTGGGGTGC 688

Qy        672 CCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATG 727
              |||| |||||||||| |||||||| ||    |||||||||| |||||||||||||||||
Db        689 CCTTTTCCTGCTGTGTTAGGGACCCAGCGGAAGATGTCCTCAATACCCAGTGTGGCTATG 748

Qy        728 ATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGG 787
              |  |||| |||||||||||||| |||||| |||||||||  ||||||||||||||||||
Db        749 ACATCCGACTCAAACTGGAGCTCGAGCAGGAGGGCTCCATCTACACCAAAGGCTGTGTGG 808

Qy        788 GCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCA 847
              |||||||||||||||||| || |||||||||||||||||||| |||||||| |||||  |
Db        809 GCCAGTTTGAGAAGTGGCTACAAGACAACCTGATCGTGGTGGCCGGGGTCTTGGTGGGGA 868

Qy        848 TCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGG 907
              |||| |||||||||| |||||  ||||| |||||||||||| ||||||||||||||||||
Db        869 TCGCCCTCCTCCAGATATTTGGCCTCTGCTTGGCCCAGAACCTCGTGAGTGACATCAAGG 928

Qy        908 CAGTGAAGGCCAACTGG 924
              | ||||||||||||||||
Db        929 CGGTGAAGGCCAACTGG 945


RESULT 4
AL543914
LOCUS       AL543914                957 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL543914 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI005YM07 5-PRIME, mRNA sequence.
ACCESSION   AL543914
VERSION     AL543914.2  GI:31265759
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 957)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12876393.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
```

```
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI005AG04QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI005AG04QP1.
FEATURES              Location/Qualifiers
     source           1. .957
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI005YM07"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      176 a    279 c    288 g    205 t      9 others
ORIGIN

  Query Match             27.3%;  Score 692.2;  DB 9;  Length 957;
  Best Local Similarity   87.9%;  Pred. No. 6.3e-83;
  Matches  813;  Conservative    3;  Mismatches   66;  Indels   43;  Gaps    4;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              ||||||:||||| |||||||||||||||||||||||||||||||||||||||| || | || |
Db         64 GCCGCASCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 123

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||||| ||| ||||||||||| ||||||
Db        124 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 183

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        184 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 243

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              |||||||| |||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        244 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 303

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
              |||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| |||
Db        304 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTG 363

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
              |||||||||| || |||||||||||||||| |||||||||||||| |||||||| |||
Db        364 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 423

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
              || ||||||||||||| ||||||||||||||||
Db        424 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTT---------------------------- 456

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
                      |||||||||||||||| ||||||| || |||||||||||||||||||||||||
Db        457 ---------CTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 507

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
              |||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||||
Db        508 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 567

Qy        551 CTCATTGACTTTGCTCAGGAATAT--TGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACT 608
              |||||||||||||||| |  |:|  |||||||||||||:||   |  |  |||||||
Db        568 CTCATTGACTTTGCTCATGTGTWTTCTGGTCTTGCTGCGGASCCGAGGCCCTCAATGACT 627

Qy        609 GGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGG 668
              |||||||||||||||| |||||||||||||||||||||| ||||| ||||||||||||
Db        628 GGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGG 687

Qy        669 TGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCT 724
              |||||||||||||||| ||||||||||||    |||||||||||||||||||||||||
Db        688 TGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCT 747

Qy        725 ATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTG 784
              | || |||||||||||||||||||||||||||||||||||| ||| |||||||||||||| |
Db        748 ACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCG 807

Qy        785 TGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGG 844
              |||||||||||||||||||||||||||||||||||||||| |||||||| || |||||  |||
Db        808 TGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGG 867

Qy        845 GCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCA 904
              ||||||| |||||||||||||||| |||||||||||||||||||| ||||||||||||||
Db        868 GCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCA 927

Qy        905 AGGCAGTGAAGGCCAACTGGATCAA 929
              |||||||||| |||||||||| |||
Db        928 AGGCAGTGAAAGCCAACTGGAGCAA 952


RESULT 5
BY708665
LOCUS       BY708665                1005 bp    mRNA    linear   EST 16-DEC-2002
DEFINITION  BY708665 RIKEN full-length enriched, adult male stomach Mus
            musculus cDNA clone 2210021G21 5', mRNA sequence.
ACCESSION   BY708665
VERSION     BY708665.1  GI:27119857
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
```

```
          ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
          Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
          Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
          Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
          Submission
           Computational Analysis of Full-Length Mouse cDNAs Compared with
          Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
           Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new
          genes. Genome Res. 10 (10), 1617-1630 (2000)
           RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res.
          10 (11), 1757-1771 (2000)
           Computer-based methods for the mouse full-length cDNA
          encyclopedia: real-time sequence clustering for construction of a
          nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
           cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
           Please visit our web site (http://genome.gsc.riken.go.jp) for
          further details.
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="2210021G21"
                     /sex="male"
                     /tissue_type="stomach"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /clone_lib="RIKEN full-length enriched, adult male
                     stomach"
                     /note="Site_1: XhoI; Site_2: SstI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and SstI."
BASE COUNT      175 a    281 c    310 g    237 t      2 others
ORIGIN

  Query Match           26.7%;  Score 677.8;  DB 14;  Length 1005;
  Best Local Similarity   88.3%;  Pred. No. 5e-81;
  Matches  760;  Conservative    0;  Mismatches   97;  Indels    4;  Gaps    2;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCCGTGGAGGGCCG 71
              ||||||||| |||||||||||||| |||| |||| || || ||||||||||  |||||| |
Db         29 GCCGCAGCCACCGGGCTAGGCCCGGGGCTGCTCCAGGCCTGGGCGGCCCGCAGAGGGCTG 88

Qy         72 ATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGC 131
                || | | ||||||| |||||||||| |||| || ||| ||||||||||||| |||||||
Db         89 GGCCGGTCACCGGCTCGGGTTCCCGGGACGGCAGGTGGCCGCTCACCATGCCCGGCAAGC 148

Qy        132 ACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 191
              ||||||| |||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        149 ACCAGCAATTCCAGGACCCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 208

Qy        192 ACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTG 251
              ||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        209 ACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATTGGCCTCTGGGCCTGGGGTG 268

Qy        252 AGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGT 311
              ||||||| |||||||||||||||||||||||||||||||| ||||||||| |||||| ||
Db        269 AGAAGGGCGTTCTCTCCAACATCTCGGCGCTGACAGATCTGGGCGGTCTTGACCCCGTGT 328

Qy        312 GGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGG 371
              ||||||||| ||||| || || |||||||| ||| ||||||| ||| ||||||| ||||
Db        329 GGCTGTTTGTGGTGGTGGGGGGAGTCATGTCAGTGTTGGGCTTCGCCCGCTGCATTGGGG 388

Qy        372 CTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCT 431
              | |||||||| ||||| |||||||||||||| ||||| |||||||| ||||||||||||
Db        389 CCCTCCGGGAAAACACCTTCCTGCTCAAGTTCTTCTCTGTGTTCCTCGGCCTCATCTTCT 448

Qy        432 TCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGC 491
              |||||||||||||  | |||||| |||||||||| |||||||| ||||| |||||||| |
Db        449 TCCTGGAGCTGGCCGCGGGGATCCTGGCCTTCGTCTTCAAGGATTGGATCCGAGACCAAC 508

Qy        492 TCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACC 551
              | ||| ||||||| |||||||| ||||| ||||| |||||||| | ||||| |||||||
Db        509 TTAATCTCTTCATCAACAACAATGTCAAAGCCTACCGGGATGATCTCGACCTTCAGAACC 568

Qy        552 TCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGA 611
              | || |||||||||||||||||| |||||||||| || |||||||||| ||||||||||
Db        569 TTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGTGCCCGAGGGCCCAATGACTGGA 628

Qy        612 ACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGC 671
              |||||||||||||| ||||||||||||||||||||||||| ||||| |||||||| |||||
Db        629 ACCTCAATATCTACTTCAACTGCACTGACTTGAACCCAAGCCGTGAGCGCTGTGGGGTCG 688

Qy        672 CCTTCTCCTGCTGTGTCAGGGAC---CCTGCGATGTCCTCAACACCCAGTGTGGCTATGA 728
              |||| |||||||||||| ||||||   |   |||||||||||| ||||||||||||||||||
Db        689 CCTTTTCCTGCTGTGTTAGGGACCAGCGGAAGATGTCCTCAATACCCAGTGTGGCTATGA 748

Qy        729 TGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGG 788
                |||| |||||||||||||| |||||| |||||||||| ||||||||||||||||||||
Db        749 CATCCGACTCAAACTGGAGCTCGAGCAGGAGGGCTCCATCTACACCAAAGGCTGTGTGGG 808

Qy        789 CCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCAT 848
              |||||||||||||||||| || |||||||||||||||||| |||||||| ||||| ||
Db        809 CCAGTTTGAGAAGTGGCTACAAGACAACCTGATCGTGGTGGCCGGGGTCTTGGTGGGGAT 868

Qy        849 CGCTCTCCTCCAGATCTTTGG 869
              ||| |||||||||| |||||
Db        869 CGC-CTCCTCCAGATATTTGG 888


RESULT 6
BU856801
LOCUS       BU856801                938 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_10475185 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6646576 5', mRNA sequence.
ACCESSION   BU856801
VERSION     BU856801.1  GI:24041791
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2885  row: j  column: 16
            High quality sequence stop: 649.
FEATURES             Location/Qualifiers
     source          1. .938
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6646576"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_107"
                     /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      162 a    290 c    279 g    207 t
ORIGIN

  Query Match             26.2%;  Score 663.8;  DB 13;  Length 938;
  Best Local Similarity   91.4%;  Pred. No. 3.8e-79;
  Matches  738;  Conservative    0;  Mismatches   62;  Indels    7;  Gaps    3;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              ||||||||||||| |||||||||||||||||||||||||||||||||||| || | || |
Db          5 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 64

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||| ||| |||||||||||| ||||||
Db         65 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 124

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        125 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 184

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        185 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 244

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
              |||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| |||
Db        245 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTG 304

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
              |||||||||| || |||||||||||||||| |||||||||||||| |||||||| |||
Db        305 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 364

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
              || |||||||||||||| |||||||| || ||||||||||
Db        365 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 424

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
              |||||||||||||||||||||||| ||||||| || ||||||||||||||||||||||||
Db        425 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 484

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
              |||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||||
Db        485 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 544

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
              |||||||||||||||||||||||| |||||||||||||||||||||| || |||||||||
Db        545 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGG 604

Qy        611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
              |||||||||||||| |||||||||||||||||||||||| ||||| ||||||||||||||
Db        605 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 664

Qy        671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
              |||||||||||||| |||||||||||||    |||||||   ||||||||||||||||
Db        665 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCCTCACACCCAGTGTGGCTAC 724

Qy        727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTG 786
              || |||||||||||||||||||||||||||||||| |||| ||| |||||||||||||| | |
Db        725 GACGTCCGGCTCAAACTGGAGCTGGAGCAGCAAGGCTTCATCCACACCAAAGGCTGCGGG 784

Qy        787 GGCC--AGTTTGAGAAGTGGCTGCAGG 811
              ||||  | ||  | ||||||||||| |
Db        785 GGCCCAATTTGAAAAAGTGGCTGCAAG 811


RESULT 7
BE615772
LOCUS       BE615772                849 bp    mRNA    linear   EST 24-AUG-2000
DEFINITION  601279927F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3621871 5',
            mRNA sequence.
ACCESSION   BE615772
VERSION     BE615772.1  GI:9897371
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM296  row: e  column: 08
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .849
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3621871"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_39"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies)."
BASE COUNT      165 a    241 c    243 g    199 t      1 others
ORIGIN

  Query Match             25.8%;  Score 655.2;  DB 10;  Length 849;
  Best Local Similarity   89.7%;  Pred. No. 5.5e-78;
  Matches  726;  Conservative    0;  Mismatches   79;  Indels    4;  Gaps    2;

Qy         87 CCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGCACCAGCACTTCCAGG 146
              |||||| |||||||||||||| ||| |||||||||||| |||||||||||||| |||||||
Db          1 CCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAGCACCAGCATTTCCAGG 60

Qy        147 AACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGG 206
              |||| ||||||||||||||||||||  |||||||||||||||||||||||||| |||||||
Db         61 AACCTGAGGTCGGCTGCTGCGGGAAA-NCTTCCTGTTTGGCTTCAACATTGTCTTCTGGG 119

Qy        207 TGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCT 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 TGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGTGAGAAGGGCGTTCTCT 179

Qy        267 CCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGG 326
              | |||||||| |||||||| |||||||||||||| ||||| |||||||||||||| || |
Db        180 CGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTGTGGCTGTTTGTGGTAG 239
```

```
Qy        327 TTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACA 386
              |||||||||||||||| |||||||||||||| |||||||| ||||| ||||||||||||||
Db        240 TTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGGGCCCTCCGGGAGAACA 299

Qy        387 CTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAA 446
              | |||||||||||||||||| |||||||| || |||||||||||||||||||||||||||
Db        300 CCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTCTTCCTGGAGCTGGCAA 359

Qy        447 CAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTA 506
              ||||||||| ||||||| || |||||||||||||||||||||||||  ||||||| |
Db        360 CAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAGCTCAACCTCTTCATCA 419

Qy        507 ACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTC 566
              |||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Db        420 ACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAACCTCATTGACTTTGCTC 479

Qy        567 AGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATT 626
              ||||||| |||||||||||||||||||| || |||||||||||||||||||||||||| |
Db        480 AGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGGAACCTCAATATCTACT 539

Qy        627 TCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTG 686
              |||||||||||||||||||| ||||| |||||||||||||||||||||||||||||||| |
Db        540 TCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTGCCCTTCTCCTGCTGCG 599

Qy        687 TCAGGGACCCTGCGATG---TCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACT 743
              |||||||||||||  |     |  | ||||||||||||| || ||||||||||||||
Db        600 TCAGGGACCCTGCGGAGGATGTCCTCAAACCCAGTGTGGCTACGACGTCCGGCTCAAACT 659

Qy        744 GGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTG 803
              |||||||||||||||||| ||| ||||| |||||||| ||||||||| |||||||||||
Db        660 GGAGCTGGAGCAGCAGGGCTTCATCCACACAAAAGGCTGCGTGGGCCAGTCTGAGAAGTG 719

Qy        804 GCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGAT 863
              ||||   |  |||||||| |||||||| ||  ||||  |||||||  |   |||||||||
Db        720 GCTGGCAGGAAACCTGATTGTGGTGGCAGGAATCTTCATGGGCATTGGCATCCTCCAGAT 779

Qy        864 CTTTGGTATCTGCCTGGCCCAGAACCTTG 892
              | | || || | |||||||| | |||  ||
Db        780 CCTCGGGATATTCCTGGCCAAAAACTCTG 808


RESULT 8
BM563474
LOCUS       BM563474               1032 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6589677 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441687
            5', mRNA sequence.
ACCESSION   BM563474
VERSION     BM563474.1  GI:18810427
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1916  row: f  column: 24
            High quality sequence stop: 697.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5441687"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      203 a    309 c    286 g    232 t      2 others
ORIGIN

  Query Match             24.9%;  Score 632.2;  DB 12;  Length 1032;
  Best Local Similarity   82.5%;  Pred. No. 5.6e-75;
  Matches  804;  Conservative    0;  Mismatches   73;  Indels   98;  Gaps    3;

Qy         80 CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGCACCAGCAC 139
              |||||||||||||| |||||||||||||| ||| |||||||||||| ||||||||||||||
Db         17 CCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAGCACCAGCAT 76

Qy        140 TTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTT 199
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTC 136

Qy        200 TTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGT 259
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        137 TTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGTGAGAAGGGC 196

Qy        260 GTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTT 319
              |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||||||||||
Db        197 GTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTGTGGCTGTTT 256

Qy        320 GTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGG 379
              || || |||||||||||||||||| |||||||||||||| |||||||| ||||| ||||||
Db        257 GTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGGGCCCTCCGG 316

Qy        380 GAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAG 439
              |||||||| ||||||||||||||||
Db        317 GAGAACACCTTCCTGCTCAAGTTT---------------------------------- 340

Qy        440 CTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTC 499
                                                  |||||||||||||||||||||||  ||
Db        341 ------------------------------GACTGGATTCGAGACCAGCTCAACCTC 367

Qy        500 TTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGAC 559
              ||||| |||||||||||||||||||| ||||| |||||||||||||||||||||||||||
Db        368 TTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAACCTCATTGAC 427

Qy        560 TTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAAT 619
              |||||||||||||| |||||||||||||||||||||||| || ||||||||||||||||||
Db        428 TTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGGAACCTCAAT 487

Qy        620 ATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCC 679
              ||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||||||
Db        488 ATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTGCCCTTCTCC 547

Qy        680 TGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATGTCCGG 735
              ||||| |||||||||||||    |||||||||||||||||||||||||||||| || ||||||
Db        548 TGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTACGACGTCCGG 607

Qy        736 CTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTT 795
              |||||||||||||||||||||||||||| ||| |||||||||||||| ||||||||||||
Db        608 CTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTGGGCCAGTTT 667

Qy        796 GAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTC 855
              |||||||||||||||||||||||||| |||||||| || |||||  |||||||||| |||
Db        668 GAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGCATCGCCCTC 727

Qy        856 CTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAG 915
              |||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||
```

```
Db        728 CTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAAGGCAGTGAAA 787

Qy        916 GCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAAACAAAACCTTGAAAA 975
              |||||||||| ||||    ||||||                          || |||||||
Db        788 GCCAACTGGAGCAAATGGAATGATG------------------------ACTTTGAAAA 822

Qy        976 CCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGGGCCTCAGCCGTGCCG 1035
              |||||||||||||||||||| |  || || |  |    | || |||||||| |    |
Db        823 CCACTGGCTTACGCCCACCATTTTCGAAGTCCTGTCCACGGCGGGGCCTCAACAAAACTC 882

Qy       1036 TCTGCCTGGGGCCCC 1050
              |||| ||||||||||
Db        883 TCTGACTGGGGCCCC 897


RESULT 9
AL556793
LOCUS       AL556793               1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL556793 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DK005YH21 5-PRIME, mRNA sequence.
ACCESSION   AL556793
VERSION     AL556793.2  GI:31278594
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12899797.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DK005CD11QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DK005CD11QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DK005YH21"
                     /cell_type="HELA CELLS COT 25-NORMALIZED"
                     /cell_line="HELA"
                     /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      231 a    322 c    324 g    258 t     66 others
ORIGIN

  Query Match             24.5%;  Score 622;  DB 9;  Length 1201;
  Best Local Similarity   90.9%;  Pred. No. 1.2e-73;
  Matches  691;  Conservative    4;  Mismatches   59;  Indels    6;  Gaps    3;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||||||||||| || | || |
Db        160 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 219

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||||| ||| |||||||||||| ||||||
Db        220 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 279

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||| |||||||||||||||||||||||||||||||||||||||
Db        280 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 339

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              |||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        340 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 399

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
              |||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| |||
Db        400 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTG 459

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
              |||||||||||| || |||||||||||||||| |||||||||||||| |||||||| |||
Db        460 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 519

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
              || |||||||||||||| |||||||||||||||||||| |||||||| || |||||||||
Db        520 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 579

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
              |||||||||||||||||||||| ||||||| || ||||||||||||||||||||||||||
Db        580 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 639

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
              |||||  ||||||| |||||||||||||||||||||||||| ||||| ||||||||||||||
Db        640 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 699

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
              |||||||||||||||||||||| |||||||||||||| :|||||| || |||||||||
Db        700 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGA-SCCGAGGCCCCAATGACTGG 758

Qy        611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGGTG 670
              |||||||||||||| |||||||||||||||||||||| ||||| ||||||||||||||||
Db        759 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGGTG 818

Qy        671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
              |||||||||||||| ||||||||||||||    |||:|||||||||||||||||||||||
Db        819 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATKTCCTCAACACCCAGTGTGGCTAC 878

Qy        727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCA 766
              || |||||||||||||||  |   | | ||| :| :  ||
Db        879 GACGTCCGGCTCAAACTGGTGAGAGGGGAGCTKGASAGCA 918


RESULT 10
BG477727
LOCUS       BG477727               1028 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602521422F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4639997 5',
            mRNA sequence.
ACCESSION   BG477727
VERSION     BG477727.1  GI:13410006
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1028)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1404  row: k  column: 06
            High quality sequence stop: 734.
FEATURES             Location/Qualifiers
     source          1. .1028
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4639997"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_20"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      194 a    296 c    310 g    228 t
ORIGIN

  Query Match             24.5%;  Score 621;  DB 10;  Length 1028;
  Best Local Similarity   86.2%;  Pred. No. 1.7e-73;
  Matches  780;  Conservative    0;  Mismatches  115;  Indels   10;  Gaps    8;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||| |||||||   ||  |   |
Db         28 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCA-GGCGGCCGCGGGGCGCTGG 86

Qy         72 ATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGC 131
                |  |  |||||||||||||| |||||||||||| ||| ||||||||||||| |||||||
Db         87 GCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAGC 146

Qy        132 ACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 191
              ||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        147 ACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 206

Qy        192 ACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTG 251
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 ACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGTG 266

Qy        252 AGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGT 311
              ||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||
Db        267 AGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTGT 326

Qy        312 GGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGG 371
              |||||||||| || |||||||||||||||| |||||||||||||| |||||||| ||||
Db        327 GGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGGG 386

Qy        372 CTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCT 431
              | |||||||||||||| |||||||||||||||||| |||||||| || ||||||||||
Db        387 CCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTCT 446

Qy        432 TCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGC 491
              |||||||||||||||||||||| ||||||| || ||||||||||||||||||||||||||
Db        447 TCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAGC 506

Qy        492 TCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACC 551
              ||||  ||||||| |||||||||||||||||||| || || ||||||||||||||||||||
Db        507 TCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGCGACGACATTGACCTCCAGAACC 566

Qy        552 TCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGA 611
              |||||||||||||||||||||| |||||||||| ||||||||| |||| ||||||||||
Db        567 TCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGA-GGCCCAATGACTGGA 625

Qy        612 ACCTCAATATCTATTTCAAC-TGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
              |||||||||||| |||||| |||||||||||||||| ||||| ||||||||||||||||
Db        626 ACCTCAATATCTACTTCAACTTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 685

Qy        671 CCCTTCTCCTGCTGTGTCAGGGAC---CCTGCGATGTCCTCAACACCCAGTGT-GGCTAT 726
              |||||||||||| || |||  |    |||||||||| ||||||||| |||||
Db        686 GCCTTCTCCTGCTGCGTAAGGACCTGGCGGAGGATGTCCTCAAAACCCAGTGTGGGCTAC 745

Qy        727 GATGTCCGGCTCAAACT-GGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGT 785
              | |||||||||||||| |||||||||||||||||| | ||| |||||||||| || ||
Db        746 AACGTCCGGCTCAAACTGGGAGCTGGAGCAGCAGGGGTTCATCCACACCAAAGG-TGCGT 804

Qy        786 GGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGG 845
              |||||||||  | |   |  |||||||||| |||||| || ||||       | |   ||||
Db        805 GGGCCAGTTGAAAAGTGGCTTGCAGGACACCCTGATTGT-GTGGGGAATCTTTCATTGGG 863

Qy        846 CATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAA 905
              ||||  |  ||| |||| ||||| ||||| | || |   ||| | || ||  |||  |
Db        864 AATCGACCCTCTCAAGATTTTTGGGATCTGGCAGGGCACAAACTTCGTAAGGAACACAAG 923

Qy        906 GGCAG 910
              |  ||
Db        924 GAGAG 928


RESULT 11
CA454987
LOCUS       CA454987                901 bp    mRNA    linear   EST 12-NOV-2002
DEFINITION  AGENCOURT_10735735 MAPcL Homo sapiens cDNA clone IMAGE:6722499 5',
            mRNA sequence.
ACCESSION   CA454987
VERSION     CA454987.1  GI:24905277
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 901)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Kristi A. Egland, Ira Pastan
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14285  row: f  column: 03
            High quality sequence stop: 683.
FEATURES             Location/Qualifiers
     source          1. .901
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6722499"
                     /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HME1
                     , LNCaP"
                     /lab_host="EMDH10B"
                     /clone_lib="MAPcL"
                     /note="Vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
                     Subtracted with brain, liver, lung, kidney and muscle.
                     Directionally cloned. Priming method: oligo-dT. Average
                     insert size: 1800 bp. Library amplification:  26,000 fold.
                     Kristi A. Egland, James J. Vincent, Robert Strausberg,
                     Bungkook Lee  & Ira Pastan:  Discovery of new breast
                     cancer genes encoding  membrane and secreted proteins.
                     Manuscript submitted."
BASE COUNT      157 a    275 c    269 g    200 t
ORIGIN

  Query Match             24.4%;  Score 619.2;  DB 14;  Length 901;
  Best Local Similarity   92.1%;  Pred. No. 3.2e-73;
  Matches  687;  Conservative    0;  Mismatches   53;  Indels    6;  Gaps    3;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||||||||||||||| || | || |
Db         25 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 84

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                 || ||| |||||||||||||| |||||||||||||| ||| |||||||||||| ||||||
Db         85 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 144
```

```
Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        145 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 204

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              |||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        205 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 264

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
              |||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| |||
Db        265 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTG 324

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
              |||||||||| || |||||||||||||||||| |||||||||||||| |||||||| |||
Db        325 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 384

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
              || |||||||||||||| |||||||||||||||||||| |||||||| || |||||||||
Db        385 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 444

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
              |||||||||||||||||||||||| ||||||| || ||||||||||||||||||||||||
Db        445 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 504

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
              |||||  ||||||| |||||||||| |||||||| ||||| ||||||||||||||||||
Db        505 CTCAACCTCTTCATCAACAACAACGACAAGGCCTACCGGGACGACATTGACCTCCAGAAC 564

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
              |||||||||||||||||||||| |||||||||||||||||||||| || |||||||||
Db        565 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGG 624

Qy        611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
              |||||||||||||| |||||||||||||||||||| ||||| ||||||||||||||||
Db        625 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 684

Qy        671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
              ||||||||||||| ||||||||||||||    |||||||||||| ||||||||||||
Db        685 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACAACCAGTGTGGCTAC 744

Qy        727 GATGTCCG-GCTCAAACTGGAGCTGG 751
                | ||||| |||| ||||||||||||
Db        745 AACGTCCGAGCTCCAACTGGAGCTGG 770
```

```
RESULT 12
BX447619
LOCUS       BX447619                888 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX447619 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG006YC20 5-PRIME, mRNA sequence.
ACCESSION   BX447619
VERSION     BX447619.1  GI:31025857
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AG006BB10QP1&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AG006BB10QP1.
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG006YC20"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      148 a    238 c    263 g    199 t     40 others
ORIGIN
```

```
  Query Match             24.1%;  Score 611.6;  DB 13;  Length 888;
  Best Local Similarity   86.7%;  Pred. No. 3.3e-72;
  Matches  664;  Conservative   35;  Mismatches   60;  Indels    7;  Gaps    3;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              ||::||:||||| |||||||||||||||||||||||:||||:||||:|| :| | || |
Db        125 GCYKCAKCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCYCAGGKCGGCYCGCKGGGCGCTG 184

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||: :|||||||||||| :|||||:|||||| :|| |||||:|||:|  ||||||
Db        185 GGCCTGGYTYCCGGCTCCGGTTTYCGGGCYGGCGGGTKGCCGCTCAMCATKCTCGGCAAG 244

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              ||||| || ||::|||||:: |||||:||||||||||||||||:|| :|||||||:||
Db        245 CACCAACATTTYYAGGAAMYTGAGGTYGGCTGCTGCGGGAAATAMTTTYTGTTTGGYTTT 304

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              ||:||| ||||:|:::| |||||||| ||:||::|||| || ||||||||:||||||||
Db        305 AAHATTTTTTTYTKKKTTCTGGGAGCTCTKTTYYTGGCTATTGGCCTCTKGGCCTGGGGT 364

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
              |||||||| |||||||: ||:|||||:  ||||||| |||||||||||||| ||||| |||
Db        365 GAGAAGGGCGTTCTCTYGAAMATCTYAACGCTGACAGATCTGGGAGGCCTTGACCCCGTG 424

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
              |||||||||| || |||||||||||||||||| |||||||||||||| |||||||| |||
Db        425 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 484

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
              || |||||||||||||| |||||||||||||||||||| |||||||| || |||||||||
Db        485 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 544

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
              |||||||||||||||||||||||| ||||||| || ||||||||||||||||||||||||
Db        545 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 604

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
              |||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||||
Db        605 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 664

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
              |||||||||||||||||||||| |||||||||||||||||||    |||| |||||||||
Db        665 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCC--GAGGCCCAATGACTGG 722

Qy        611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
              |||||||||||||| |||||||||||||||||||| ||||| |||||||||||||||||
Db        723 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 782

Qy        671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
              ||||||||||||| |||||||||||||    |||||:||||||||||||||||||||
Db        783 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTYCTCAACACCCAGTGTGGCTAC 842

Qy        727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACA 772
              || |||||||||||||||||||||||||||||||||| ||| ||||
Db        843 GACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACA 888
```

```
RESULT 13
BG770931
LOCUS       BG770931                781 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602719148F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4839779 5',
            mRNA sequence.
ACCESSION   BG770931
VERSION     BG770931.1  GI:14081584
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1669  row: o  column: 12
            High quality sequence stop: 697.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4839779"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_60"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      127 a    239 c    239 g    176 t
ORIGIN

  Query Match             23.9%;  Score 606.4;  DB 12;  Length 781;
  Best Local Similarity   89.4%;  Pred. No. 1.8e-71;
  Matches  676;  Conservative    0;  Mismatches   76;  Indels    4;  Gaps    2;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||||||  | ||   ||  |   |
Db         22 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGCGGCCCGCGGGGCGCTGG 81

Qy         72 ATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGC 131
                |  |  |||||||||||| |||||||||||| ||| |||||||||||| |||||||
Db         82 GCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAGC 141

Qy        132 ACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 191
              ||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        142 ACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 201

Qy        192 ACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTG 251
              ||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||
Db        202 ACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGTG 261

Qy        252 AGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGT 311
              ||||||| |||||||| |||||||| |||||||| |||||||||||||| ||||| ||||
Db        262 AGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTGT 321

Qy        312 GGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGG 371
              |||||||||| || |||||||||||||||| |||||||||||||| |||||||| ||||
Db        322 GGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGGG 381

Qy        372 CTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCT 431
              | |||||||||||||| |||||||||||||||||||| |||||||| || ||||||||||
Db        382 CCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTCT 441

Qy        432 TCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGC 491
              |||||||||||||||||||||| ||||||| || ||||||||||||||||||||||||||
Db        442 TCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAGC 501

Qy        492 TCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACC 551
              ||||  ||||||| |||||||||||||||||||| ||||| ||||||||||||||||||
Db        502 TCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAACC 561

Qy        552 TCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGA 611
              |||||||||||||||||||||| |||||||||||||||||||||| || ||||||||||
Db        562 TCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGGA 621

Qy        612 ACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGC 671
              |||||||||||| |||||||||||||||||||| ||||| ||||||||||||||||||
Db        622 ACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTGC 681

Qy        672 CCTTCTCCTGCTGTGTCAGGGACCCTGCG---ATGTCCTCAACACCCAGTGTGGCTATGA 728
              |||||||||||| |||   |||||||||     |     | |||||||||||||| ||
Db        682 CCTTCTCCTGCTGCGTCCAAGACCCTGCGGAAGGATGTCTCAAACCCAGTGTGGCTACGA 741

Qy        729 TGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTC 764
              ||||||||||||||||||||    |||| ||
Db        742 CGTCCGG-TCAAACTGGAGCTGGAGGCAGAGGGTTC 776


RESULT 14
BI914325
LOCUS       BI914325                718 bp    mRNA    linear   EST 17-OCT-2001
DEFINITION  603182305F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246304 5',
            mRNA sequence.
ACCESSION   BI914325
VERSION     BI914325.1  GI:16198834
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11621  row: j  column: 01
            High quality sequence stop: 712.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5246304"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_121"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
```

Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."

BASE COUNT 115 a 222 c 217 g 164 t
ORIGIN

Query Match 23.7%; Score 602.4; DB 12; Length 718;
Best Local Similarity 91.3%; Pred. No. 6.3e-71;
Matches 639; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy          1 CCACGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCC 60
              | | || |   |||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 CGAGGCATGAAGCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCG 60

Qy         61 GTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCAT 120
                ||  |   |  |  | |||||||||||||| |||||||||||||| ||| |||||||||
Db         61 CGGGGCGCTGGGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCAT 120

Qy        121 GCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCT 180
              ||| |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||
Db        121 GCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCT 180

Qy        181 GTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTG 240
              |||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db        181 GTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTG 240

Qy        241 GGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCT 300
              |||||||||||||||||| |||||||| |||||||| |||||||| ||||||||||||||
Db        241 GGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCT 300

Qy        301 CGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGG 360
              |||||| |||||||||||||| || |||||||||||||||| |||||||||||||| ||
Db        301 TGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGG 360

Qy        361 CTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGG 420
              |||||| ||||| |||||||||||||| |||||||||||||||||| |||||||| ||
Db        361 CTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGG 420

Qy        421 CCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGAT 480
              |||||||||||||||||||||||||||||||||| ||||||| || ||||||||||||||
Db        421 TCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGAT 480

Qy        481 TCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGA 540
              |||||||||||||||  ||||||| |||||||||||||||||||| ||||| ||||||||
Db        481 TCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGA 540

Qy        541 CCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCC 600
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db        541 CCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGCCC 600

Qy        601 TAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCG 660
              |||||||||||||||||||||||| |||||||||||||||||||||||| ||||| ||||
Db        601 CAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCG 660

Qy        661 CTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCG 700
              |||||||||||||||||||||||| ||||||||||||||
Db        661 CTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCG 700
```

RESULT 15
BG478644
LOCUS BG478644 902 bp mRNA linear EST 21-MAR-2001
DEFINITION 602525532F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4643959 5', mRNA sequence.
ACCESSION BG478644
VERSION BG478644.1 GI:13410923
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1414 row: p column: 08
High quality sequence stop: 813.
FEATURES Location/Qualifiers
source 1. .902
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4643959"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_20"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 146 a 268 c 288 g 200 t
ORIGIN

Query Match 22.4%; Score 568.4; DB 10; Length 902;
Best Local Similarity 88.2%; Pred. No. 1.8e-66;
Matches 711; Conservative 0; Mismatches 81; Indels 14; Gaps 8;

```
Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||||||||||||| || | || |
Db         28 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 87

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||| ||| |||||||||||| ||||||
Db         88 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 147

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||| |||||||||||||||||||||||||||||||||||||||
Db        148 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 207

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
              |||||||| |||||||||||||||||||||||||||| |||||||||||||||||||||
Db        208 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 267

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGAT-CTGGGAGGCCTCGACCCTGT 309
              ||||||| |||||||| |||||||| ||| |||||||||| |||||||||||||| ||
Db        268 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCCTGGGAGGCCTTGACCCCGT 327

Qy        310 GTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGG 369
              |||||||||||| || |||||||||||||||| |||||||||||||| |||||||| ||
Db        328 GTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGG 387

Qy        370 GGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTT 429
              ||| |||||||||||||| |||||||||||||||||||| |||||||| || ||||||||
Db        388 GGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTT 447

Qy        430 CTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCA 489
              |||||||||||||||||||||||||| ||||||| || ||||||||||||||||||||||
```

```
Db        448 CTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCA 507

Qy        490 GCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGA-TGACATTGACCTCCAGA 548
              ||||||  ||||||| ||||||||||||||||||||| |||||  ||||||||||||||||
Db        508 GCTCAAACTCTTCATCAACAACAACGTCAAGGCCTACCGGGACCGACATTGACCTCCAGA 567

Qy        549 ACCTCATTGACTTTG-CTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGAC 607
              ||||||||||||||| |||||||||| |||||||||| ||||||||| |||| ||||||
Db        568 ACCTCATTGACTTTGCCTCAGGAATACTGGTCTTGCTGTGGAGCCCGA-GGCCCAATGAC 626

Qy        608 TGGAACCTCAATATCTATTTCAACTGC-ACTGACTTGAACCCGAGCCGAGAGCGCTGC-G 665
              ||||||||||||||| ||||||||| |||||||||  || ||||| ||||||||| |
Db        627 TGGAACCTCAATATCTACTTCAACTGCAACTGACTTGAGACCCAGCCGGGAGCGCTGCAG 686

Qy        666 GGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTCAACACCCAGTG------ 719
              |||||||||||||||||| ||||||||||||||      |   |  | || |
Db        687 GGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGGAGGATGTCCCTCGAGAGCCAGTG 746

Qy        720 -TGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778
               ||||||       | | |  |||||||||||||||||||||||| ||| |    |||||
Db        747 TTGGCTACCGACGTCCGGTTCAACTGGAGCTGGAGCAGCAGGGCTTCATCCCAGACAAAG 806

Qy        779 GCTGTGTGGGCCAGTTTGAGAAGTGG 804
              |||| |||||||||| |||||||||
Db        807 GCTGCGTGGGCCAGTTGGAGAAGTGG 832

Search completed: November 21, 2003, 18:57:42
Job time : 5188 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 15:34:38 ; Search time 654 Seconds
(without alignments)
10475.801 Million cell updates/sec

Title: US-09-972-970-2
Perfect score: 2538
Sequence: 1 ccacgcgtccggccgcagcc.....aaaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2538	100.0	2538	24	AAD21883	Human TM4SF recept
2	1019	40.1	2091	21	AAA96482	cDNA encoding a hu
3	948.4	37.4	1644	24	ABL90838	Human polynucleoti
4	790.4	31.1	3185	24	AAI72633	Human 23228 cDNA.
5	787.2	31.0	1126	22	AAC90014	Clone HNTMH27 codi
6	773.4	30.5	1023	24	ABZ11816	Human polynucleoti
7	511	20.1	1567	23	ACC44092	Rat gamma-hydroxyb
8	449.2	17.7	1735	22	AAH98169	Human EST-derived
9	441.6	17.4	1655	21	AAC98195	Human colon cancer
10	418.6	16.5	1174	22	AAC90015	Clone HE8EJ16 codi
11	402.2	15.8	1178	22	AAC90020	Clone HE8EL16 codi
12	370	14.6	1988	24	ABK35735	cDNA sequence #126
13	357.4	14.1	2672	21	AAZ65259	Human secreted pro
14	354.8	14.0	2465	25	ABX70953	Novel human cDNA s
15	353.4	13.9	813	22	AAF90629	Human TANGO 339 cD
16	353.4	13.9	2715	22	AAF90628	Human TANGO 339 cD
17	351.8	13.9	813	22	AAF90656	Human TANGO 339 K1
18	351.8	13.9	813	22	AAF90657	Human TANGO 339 F2
19	351.8	13.9	813	22	AAF90658	Human TANGO 339 A4
20	351.8	13.9	813	22	AAF90659	Human TANGO 339 D6
21	351.8	13.9	1388	24	AAD29623	Human tetraspan pr
22	351.8	13.9	2623	23	AAS92016	DNA encoding novel
23	343.8	13.5	1428	21	AAA93623	Human tetraspanin-
24	333.6	13.1	1110	20	AAZ20854	Polynucleotide seq
25	333.6	13.1	1110	22	AAS59291	Human cDNA encodin
26	333.6	13.1	1110	24	ABA90960	Human polynucleoti
27	321	12.6	1667	21	AAZ65350	Human secreted pro
28	295.8	11.7	864	24	AAI72287	NET-4 antisense mo
29	294.2	11.6	368	21	AAZ93364	Sequence encoding
30	294.2	11.6	368	24	AAL41055	cDNA of Human F-bo
31	254.4	10.0	816	21	AAA93675	Human tetraspanin-
32	244.6	9.6	394	22	AAF67433	Novel human polynu
33	241.8	9.5	601	23	AAS80667	DNA encoding novel
34	240	9.5	839	22	AAH34927	Human colon cancer
35	234	9.2	1932	21	AAZ98047	Human secreted pro
36	234	9.2	1932	22	AAD11660	Human secreted pro
37	234	9.2	1932	24	ABK69756	Human secreted pro
38	234	9.2	2384	24	ABZ11836	Human polynucleoti
39	233.2	9.2	852	22	AAS44553	Human cDNA encodin
40	233	9.2	389	22	AAF67421	Novel human polynu
41	225	8.9	2638	24	ABL90058	Human polynucleoti
42	174.2	6.9	353	21	AAZ42460	Human 5' EST isola
43	166.6	6.6	1120	23	ABL16799	Drosophila melanog
44	158.2	6.2	810	23	ABL14835	Drosophila melanog
45	156	6.1	1214	23	AAS80668	DNA encoding novel

ALIGNMENTS

RESULT 1
AAD21883
ID AAD21883 standard; cDNA; 2538 BP.
XX
AC AAD21883;
XX
DT 12-FEB-2002 (first entry)
XX
DE Human TM4SF receptor-encoding gene 1 cDNA clone HOFOB55, SEQ ID NO:2.
XX
KW Human; 4-transmembrane superfamily; TM4SF receptor; breast cancer;
KW ovarian cancer; immune disorder; Addison's disease; wound healing;
KW gene therapy; autoimmune haemolytic anaemia; autoimmune thyroiditis;
KW diabetes mellitus; allergy; Crohn's disease; multiple sclerosis;
KW rheumatoid arthritis; ulcerative colitis; cardiovascular disorder;
KW myocardial ischaemia; neurological disease; antifungal; antiviral;
KW antibacterial; cerebral anoxia; epilepsy; infectious disease;
KW antiparasitic; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 119..820
FT /*tag= a
FT /product= "Human TM4SF receptor protein"
XX
PN WO200177173-A1.
XX
PD 18-OCT-2001.
XX
PF 05-APR-2001; 2001WO-US11130.

XX
PR 10-APR-2000; 2000US-195336P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Shi Y, Ruben SM;
XX
DR WPI; 2002-017447/02.
DR P-PSDB; AAE13218.
XX
PT Novel isolated protein, a member of 4-transmembrane superfamily of
PT receptor polypeptides, useful in the prevention, treatment and
PT diagnosis of cancer, immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 1; Page 265-266; 271pp; English.
XX
CC The invention relates to human 4-transmembrane superfamily (TM4SF)
CC receptor polypeptides and polynucleotides. Sequences of the invention
CC are useful for preventing, treating, ameliorating or diagnosing a
CC pathological condition or a susceptibility to a pathological condition.
CC TM4SF polypeptides are useful for screening molecules which modify
CC their activity. TM4SF nucleic acids, protein, antibodies, agonists and
CC antagonists are useful in the diagnosis, treatment and prevention of
CC cancer, particularly breast and ovarian cancer, and other cancers of
CC the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital; immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. TM4SF polynucleotides are
CC also useful in gene therapy. The present sequence is human TM4SF
CC receptor protein encoding cDNA.
XX
SQ Sequence 2538 BP; 528 A; 669 C; 747 G; 594 T; 0 other;

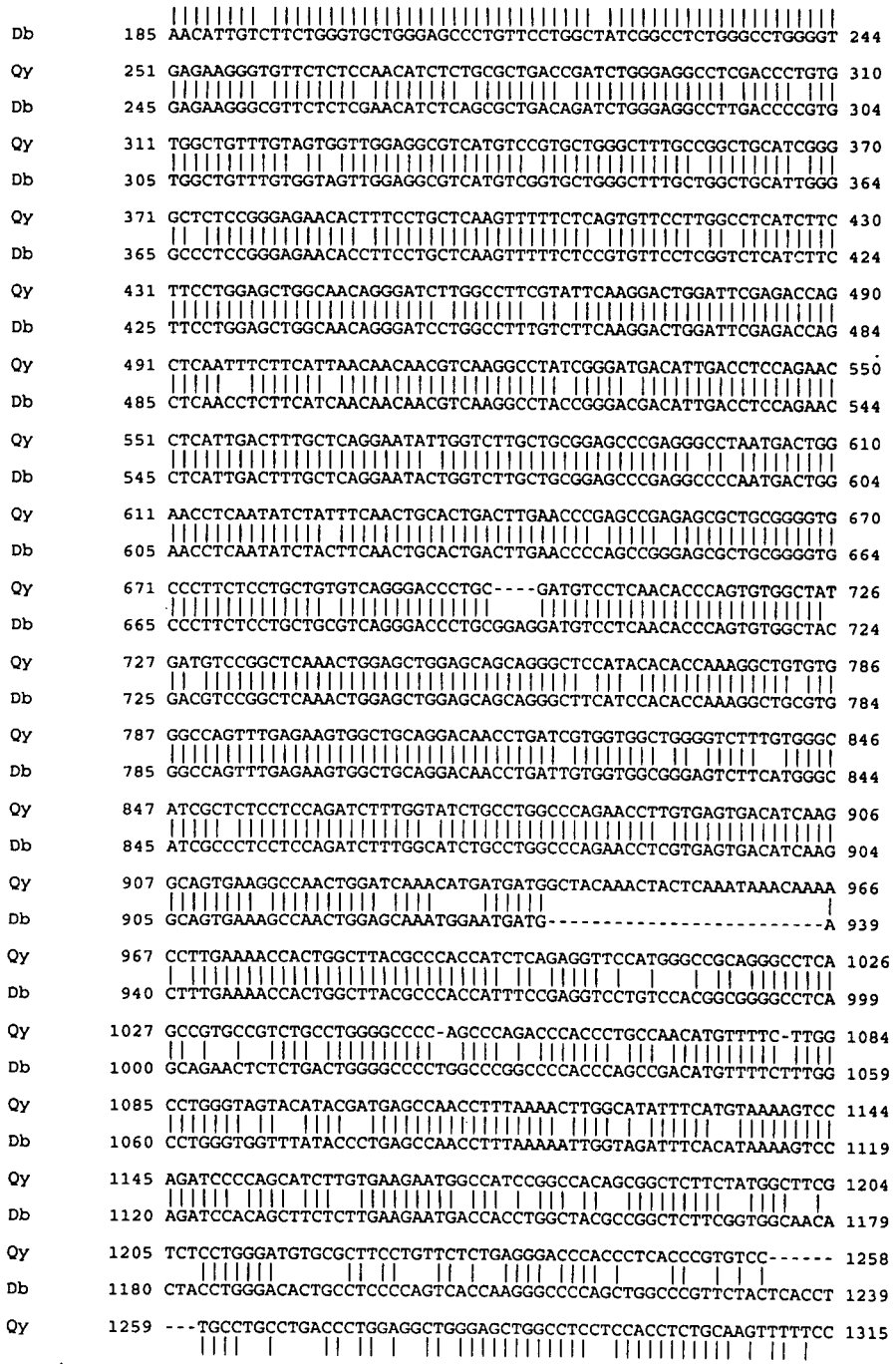
Query Match 100.0%; Score 2538; DB 24; Length 2538;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

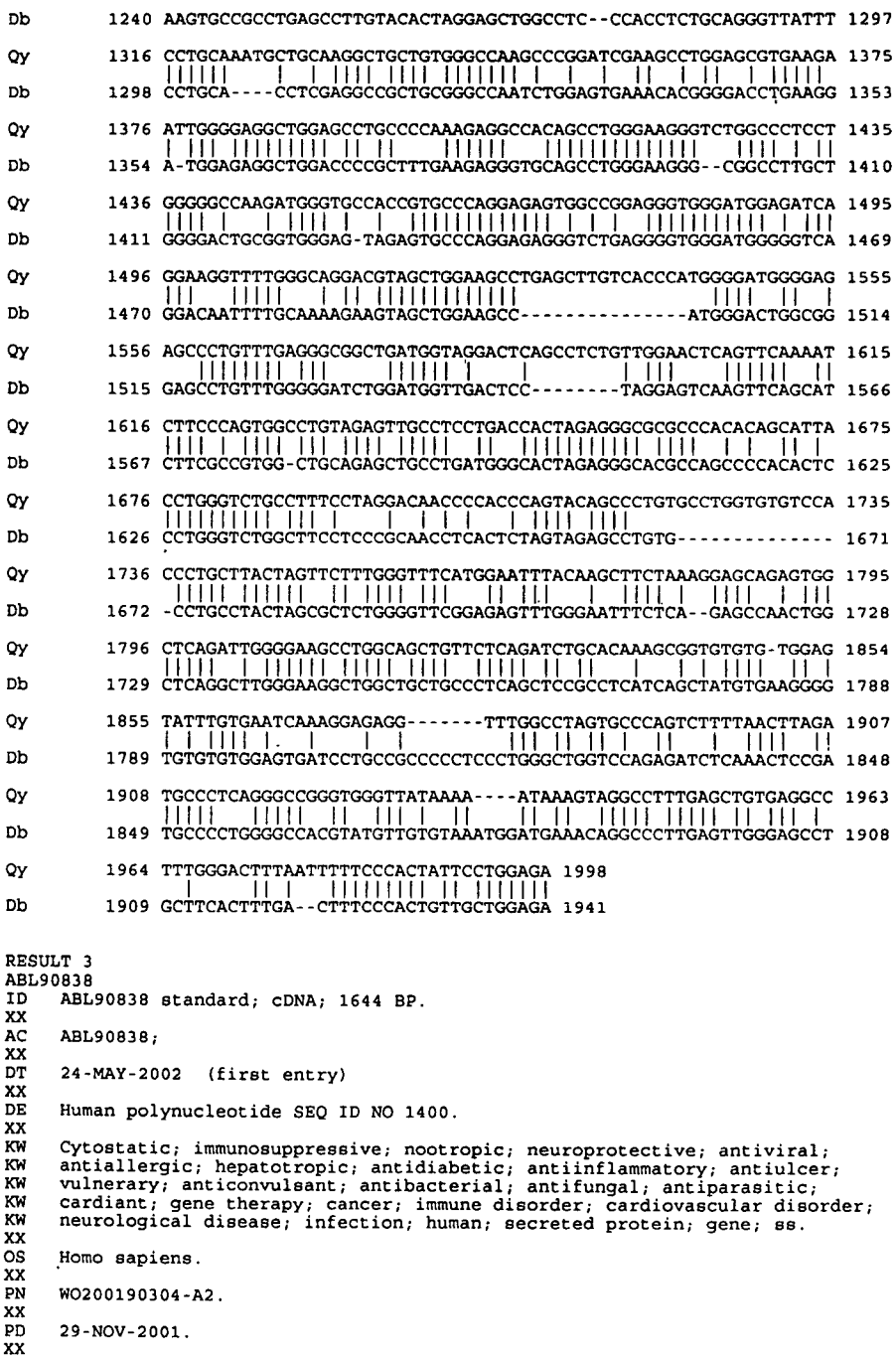
```
Qy          1 CCACGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCACGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCC 60

Qy         61 GTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCAT 120

Qy        121 GCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCT 180

Qy        181 GTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTG 240

Qy        241 GGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCT 300

Qy        301 CGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGG 360

Qy        361 CTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGG 420

Qy        421 CCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGAT 480

Qy        481 TCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGA 540

Qy        541 CCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCC 600

Qy        601 TAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCG 660

Qy        661 CTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTCAACACCCAGTGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTCAACACCCAGTGT 720

Qy        721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGC 780

Qy        781 TGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTT 840

Qy        841 GTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGAC 900

Qy        901 ATCAAGGCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATCAAGGCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAA 960

Qy        961 ACAAAACCTTGAAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACAAAACCTTGAAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGG 1020

Qy       1021 GCCTCAGCCGTGCCGTCTGCCTGGGGCCCCAGCCCAGACCCACCCTGCCAACATGTTTTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GCCTCAGCCGTGCCGTCTGCCTGGGGCCCCAGCCCAGACCCACCCTGCCAACATGTTTTC 1080

Qy       1081 TTGGCCTGGGTAGTACATACGATGAGCCAACCTTTAAAACTTGGCATATTTCATGTAAAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTGGCCTGGGTAGTACATACGATGAGCCAACCTTTAAAACTTGGCATATTTCATGTAAAA 1140

Qy       1141 GTCCAGATCCCCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GTCCAGATCCCCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGC 1200

Qy       1201 TTCGTCTCCTGGGATGTGCGCTTCCTGTTCTCTGAGGGACCCACCCTCACCCGTGTCCTG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTCGTCTCCTGGGATGTGCGCTTCCTGTTCTCTGAGGGACCCACCCTCACCCGTGTCCTG 1260

Qy       1261 CCTGCCTGACCCTGGAGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTCCCCTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCTGCCTGACCCTGGAGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTCCCCTGC 1320

Qy       1321 AAATGCTGCAAGGCTGCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAGAATTGG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAATGCTGCAAGGCTGCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAGAATTGG 1380

Qy       1381 GGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCCTGGGGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCCTGGGGG 1440

Qy       1441 CCAAGATGGGTGCCACCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATCAGGAAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCAAGATGGGTGCCACCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATCAGGAAG 1500

Qy       1501 GTTTTGGGCAGGACGTAGCTGGAAGCCTGAGCTTGTCACCCATGGGGATGGGGAGAGCCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GTTTTGGGCAGGACGTAGCTGGAAGCCTGAGCTTGTCACCCATGGGGATGGGGAGAGCCC 1560
```

```
Qy       1561 TGTTTGAGGGCGGCTGATGGTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTTCC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TGTTTGAGGGCGGCTGATGGTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTTCC 1620

Qy       1621 CAGTGGCCTGTAGAGTTGCCTCCTGACCACTAGAGGGCGCGCCCACACAGCATTACCTGG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGTGGCCTGTAGAGTTGCCTCCTGACCACTAGAGGGCGCGCCCACACAGCATTACCTGG 1680

Qy       1681 GTCTGCCTTTCCTAGGACAACCCCACCCAGTACAGCCCTGTGCCTGGTGTGTCCACCCTG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTCTGCCTTTCCTAGGACAACCCCACCCAGTACAGCCCTGTGCCTGGTGTGTCCACCCTG 1740

Qy       1741 CTTACTAGTTCTTTGGGTTTCATGGAATTTACAAGCTTCTAAAGGAGCAGAGTGGCTCAG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTTACTAGTTCTTTGGGTTTCATGGAATTTACAAGCTTCTAAAGGAGCAGAGTGGCTCAG 1800

Qy       1801 ATTGGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGTGGAGTATTTG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ATTGGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGTGGAGTATTTG 1860

Qy       1861 TGAATCAAAGGAGAGGTTTGGCCTAGTGCCCAGTCTTTTAACTTAGATGCCCTCAGGGCC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGAATCAAAGGAGAGGTTTGGCCTAGTGCCCAGTCTTTTAACTTAGATGCCCTCAGGGCC 1920

Qy       1921 GGGTGGGTTATAAAAATAAAGTAGGCCTTTGAGCTGTGAGGCCTTTGGGACTTTAATTTT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GGGTGGGTTATAAAAATAAAGTAGGCCTTTGAGCTGTGAGGCCTTTGGGACTTTAATTTT 1980

Qy       1981 TCCCACTATTCCTGGAGATGGGACATAGAGAGACATTGCTTTGTGCTGAGAAATACTTGC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TCCCACTATTCCTGGAGATGGGACATAGAGAGACATTGCTTTGTGCTGAGAAATACTTGC 2040

Qy       2041 ATGATTGAGTCTGAGTCGCTAAGGGCAACTGGCCTTGAGTGACATCAAGGGGTGGTGGGG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ATGATTGAGTCTGAGTCGCTAAGGGCAACTGGCCTTGAGTGACATCAAGGGGTGGTGGGG 2100

Qy       2101 ACTGTGGCAAACCACAGATTCCCACCTGAAATTGGTGGCTGTCCTTCCGTTGGGGCTAAT 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACTGTGGCAAACCACAGATTCCCACCTGAAATTGGTGGCTGTCCTTCCGTTGGGGCTAAT 2160

Qy       2161 GGCTGTACAGCGAGAATGTAGGTAGGTCTGTCTAATGGGAGAAGTCTGGAGAAGCCAAGA 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCTGTACAGCGAGAATGTAGGTAGGTCTGTCTAATGGGAGAAGTCTGGAGAAGCCAAGA 2220

Qy       2221 AGCTAGATTTTTCATGTGAACTATCCCGAGTTTTAAGTTGTTTGCAGCTAATGAGAAAAA 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 AGCTAGATTTTTCATGTGAACTATCCCGAGTTTTAAGTTGTTTGCAGCTAATGAGAAAAA 2280

Qy       2281 CCTCTTAAACCCTGATAGTCAAAAGGTGTGGGGGCCATCTTTGACACCTCCCCCACCATA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCTCTTAAACCCTGATAGTCAAAAGGTGTGGGGGCCATCTTTGACACCTCCCCCACCATA 2340

Qy       2341 GGTCCCTCAGGGACAGTGCCCCATGGGAGCCCTGGTGAGTCCACGGTTCAGGAATGCTGG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GGTCCCTCAGGGACAGTGCCCCATGGGAGCCCTGGTGAGTCCACGGTTCAGGAATGCTGG 2400

Qy       2401 GAACTGCTGCAGGTGGGCGGGTTGTGGGCCAGCACCCATCGTGGCTCCCAGGTGTGGGCT 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GAACTGCTGCAGGTGGGCGGGTTGTGGGCCAGCACCCATCGTGGCTCCCAGGTGTGGGCT 2460

Qy       2461 GGGCTGGGCTATGGGTGGGCTCTACATGCTACAATAAATGGGGCTCATGATAAAAAAAAA 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GGGCTGGGCTATGGGTGGGCTCTACATGCTACAATAAATGGGGCTCATGATAAAAAAAAA 2520

Qy       2521 AAAAAAAAAAAAAAAAA 2538
             |||||||||||||||||
Db       2521 AAAAAAAAAAAAAAAAA 2538
```

```
RESULT 2
AAA96482
ID   AAA96482 standard; cDNA; 2091 BP.
XX
AC   AAA96482;
XX
DT   08-FEB-2001  (first entry)
XX
DE   cDNA encoding a human transmembrane protein.
XX
KW   Human; transmembrane protein; cell proliferation disorder;  myeloma;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW   allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW   Alzheimer's disease; Tourette's disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             113..1102
FT                   /*tag=  a
XX
PN   WO200056891-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07817.
XX
PR   22-MAR-1999;   99US-0125537.
PR   16-JUN-1999;   99US-0139565.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Yue H,  Lal P,  Tang YT,  Hillman JL,  Reddy R,  Bandman O,  Baughn MR;
PI   Lu DAM,  Azimzai Y,  Yang J;
XX
DR   WPI; 2000-579485/54.
DR   P-PSDB; AAB18969.
XX
PT   New human transmembrane proteins are used to treat a disease or
PT   condition associated with decreased expression of functional HTMP e.g.
PT   Tourette's disorder, angina and leukaemia -
XX
PS   Claim 4; Page 113-114; 130pp; English.
XX
CC   The present sequence encodes a human transmembrane proteins (HTMP).
CC   Agonists and antagonists of the protein are used to treat a disease
CC   or condition associated with overexpression of the protein. Diseases
CC   and conditions which can be treated include cell proliferative,
CC   immunological, reproductive, smooth muscle and neurological disorders
CC   e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC   syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC   stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC   polynucleotides may be used to detect and quantify gene expression in
CC   biopsied tissues where protein expression may be correlated with disease
CC   e.g. to determine absence, presence or excess expression of HTMP or to
CC   monitor regulation of HTMP expression during therapeutic intervention.
XX
SQ   Sequence 2091 BP; 392 A; 602 C; 637 G; 460 T; 0 other;
```

```
  Query Match           40.1%;  Score 1019;  DB 21;  Length 2091;
  Best Local Similarity   76.6%;  Pred. No. 3.3e-236;
  Matches 1544;  Conservative    0;  Mismatches  365;  Indels  106;  Gaps   20;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
              |||||||||||| |||||||||||||||||||||||||||||||||| || | || |
Db          5 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 64

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
                || ||| |||||||||||||| |||||||||||||| ||| ||||||||||||| ||||||
Db         65 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 124

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
              |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        125 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 184

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
```





RESULT 3
ABL90838
ID ABL90838 standard; cDNA; 1644 BP.
XX
AC ABL90838;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1400.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX

PF 18-MAY-2001; 2001WO-US16450.
XX
PR 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-122018/16.
DR P-PSDB; ABB90429.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 4; SEQ ID NO 1400; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1644 BP; 297 A; 479 C; 513 G; 346 T; 9 other;

Query Match 37.4%; Score 948.4; DB 24; Length 1644;
Best Local Similarity 82.1%; Pred. No. 3.6e-219;
Matches 1244; Conservative 6; Mismatches 213; Indels 52; Gaps 12;

```
Qy        12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
             ||||||||||||| |||||||||||||||||||||||||||||||||||| || | || |
Db       114 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 173

Qy        72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
               || ||| |||||||||||||| |||||||||||||| ||| |||||||||||| ||||||
Db       174 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 233

Qy       131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
             |||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       234 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 293

Qy       191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
             |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       294 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 353

Qy       251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
             |||||||| |||||||| |||||||: |||||||| ||||||||||||| ||:|| |||
Db       354 GAGAAGGGCGTTCTCTCGAACATCTYAGCGCTGACAGATCTGGGAGGCCTTGAMCCCGTG 413

Qy       311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
             |||||||||||  || |||||||||||||||| |||||||||||||| |||||||| |||
Db       414 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 473

Qy       371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
             ||  |||||||||||||| |||||||||||||||||||||||| |||||||| || |||||||||
Db       474 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 533

Qy       431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
             |||||||||||||||||||||||||| ||||||| || ||||||||||||||||||||||
Db       534 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 593

Qy       491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
             |||||  |||||||| |||||||||||||||||||| |||||  |||||||||||||||||
Db       594 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 653

Qy       551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
             |||||||||||||||||||||||| |||||||||||||||||||||| || |||||||||
Db       654 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGG 713

Qy       611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
             |||||||||||||| |||||||||||||||||||||||| ||||| ||||||||||||||||
Db       714 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 773

Qy       671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
             |||||||||||||| |||||||||||||    |||||||||||||||||||||||||||||
Db       774 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTAC 833

Qy       727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTG 786
             || |||||||||||||||||||||||||||||||||||| ||| ||||||||||||| |||
Db       834 GACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTG 893

Qy       787 GGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGC 846
             |||||||||||||||||||||||||||||||||||| |||||||| || |||||  |||||
Db       894 GGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGC 953

Qy       847 ATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAG 906
             |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       954 ATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAAG 1013

Qy       907 GCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAAACAAAA 966
             |||||||| |||||||||| ||||     ||||||                       |
Db      1014 GCAGTGAAAGCCAACTGGAGCAAATGGAATGATG------------------------A 1048

Qy       967 CCTTGAAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGGGCCTCA 1026
             | |||||||||||||||||||||||||||||| || |||||  |   |  || ||||||||
Db      1049 CTTTGAAAACCACTGGCTTACGCCCACCATTTCCGAGGTCCTGTCCACGGCGGGGCCTCA 1108

Qy      1027 GCCGTGCCGTCTGCCTGGGGCCCC-AGCCCAGACCCACCCTGCCAACATGTTTTC-TTGG 1084
             || |  |  |||| |||||||||  |||  | |||||||  ||| |||||||||  ||||
Db      1109 GCAGAACTCTCTGACTGGGGCCCCTGGCCNGGCCCCACCCAGCCGACATGTTTTCTTTGG 1168

Qy      1085 CCTGGGTAGTACATACGATGAGCCAACCTTTAAAACTTGGCATATTTCATGTAAAAGTCC 1144
             |||||||  ||  ||||  |||||||||||||||| |||| | ||||||  |||||||||
Db      1169 CCTGGGTGGTTTATACCCTGAGCCAACCTTTAAAAATTGGTAGATTTCACATAAAAGTCC 1228

Qy      1145 AGATCCCCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGCTTCG 1204
             |||||| |||| |||  |||||||||  ||| | ||| ||  ||||||||||  ||||  |
Db      1229 AGATCCACAGCTTCTCTTGAAGAATGACCACCTGGCTAC-GCGGCTCTTCGGTGGCAACA 1287

Qy      1205 TCTCCTGGGATGTGCGCTTCCTGTTCTCTGAGGGACCCACCCTCACCCGTGTCC------ 1258
             |||||||        || ||   || |  |||| |||| |     ||  | | |
Db      1288 CTACCTGGGACACTGCCTCCCCAGTCACCAAGGGCCCCAGCTGGCCCGTTCTACTCACCT 1347

Qy      1259 ---TGCCTGCCTGACCCTGGAGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTCC 1315
                ||||  |    || || |  || ||||||||||||  ||||||||||  | || |
Db      1348 AAGTGCCGCCTGACCCTTGTACACTAGGAGCTGGCCTC--CCACCTCTGCAGGGTTATTT 1405

Qy      1316 CCTGCAAATGCTGCAAGGCTGCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAGA 1375
             ||||||    |  | |||| |||| |||||||  |  |  |  ||  | ||  | |||||
Db      1406 CCTGCA----CCTCGAGGCCGCTGCGGGCCAATCTGGAGTGAAACACGGGGACCTGAAGG 1461

Qy      1376 ATTGGGGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCCT 1435
             | ||| |||||||||  || ||    ||||||  ||||||||||||||    |||| | ||
Db      1462 A-TGGAGAGGCTGGACCCCGCTTTGAAGAGGGTGCAGCCTGGGAAGGG--CGGCCTTGCT 1518

Qy      1436 GGGGGCCAAGATGGGTGCCACCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATCA 1495
             |||| |   | |||| |  |  ||:|||||:|: | | |  ||||:||| | |
Db      1519 GGGGACTGCGGTGGGAG-TAGAGTKCCCAGGARARGGTCTGAGGGGTKGGATTGGGGGKC 1577

Qy      1496 GGAAGGTTTTGGGCA 1510
             | |   |||  |||
Db      1578 AGGACAATTTTTGCA 1592
```

```
RESULT 4
AAI72633
ID   AAI72633 standard; cDNA; 3185 BP.
XX
AC   AAI72633;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human 23228 cDNA.
XX
KW   Gene; tetraspanin; 23228; cell surface protein; transmembrane domain;
KW   extracellular growth factor; HB-EGF; TGF-alpha; amphiregulin;
KW   diabetes mellitus; arthritis; multiple sclerosis; encephalomyelitis;
KW   dermatitis; Crohn's disease; and asthma; cancer; metastasis;
KW   viral infection; cellular signalling activity; cell proliferation;
KW   cell motility; CD81; B-Cell antigen receptor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             168..980
FT                   /*tag=  a
FT                   /product= "23228"
XX
PN   WO200216603-A2.
XX
PD   28-FEB-2002.
XX
PF   21-AUG-2001; 2001WO-US41811.
XX
PR   21-AUG-2000; 2000US-226612P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Leiby KR;
XX
DR   WPI; 2002-315466/35.
DR   P-PSDB; AAB47950.
XX
PT   New tetraspanin 23228 polypeptide useful in screening assays,
PT   predictive medicine and as a prophylactic or therapeutic agent, e.g.,
PT   for hematopoietic and immune diseases such as diabetes or multiple
PT   sclerosis -
XX
PS   Claim 1; Page 111-113; 123pp; English.
XX
CC   This sequence encodes tetraspanin 23228. This protein is a cell surface
CC   protein having four transmembrane domains (TM1-TM4). TM1, TM3 and TM4
CC   have a single polar amino acid located within them, which may interact
CC   with each other and contribute to domain stability. The cytoplasmic N-
CC   and C-terminal domains and the intracellular loop between TM2 and TM3
CC   are less than 30 amino acids in length. The second extracellular domain
CC   between TM3 and TM4 contains conserved Cys residues and may function
CC   to bind extracellular growth factors, such as HB-EGF, TGF-alpha and
CC   amphiregulin. The 23228 protein is useful for diagnosing and treating
CC   23228-mediated disorders, e.g., haematopoietic and/or immune disorders
CC   such as diabetes mellitus, arthritis, multiple sclerosis,
CC   encephalomyelitis, dermatitis, Crohn's disease and asthma; cell
CC   proliferation and differentiation disorders, e.g., cancers or
CC   metastasis; and/or viral infections. 23228 polypeptide can control
CC   cellular signalling activity, bind to an extracellular growth factor,
CC   for example, amphiregulin, regulate cell proliferation, bind to a cell
CC   surface protein, to recruit intracellular kinases, to regulate cell
CC   motility, bind to another tetraspanin such as CD81, to associate with a
CC   B-Cell antigen receptor and the ability to modulate the association
CC   with a virus.
XX
SQ   Sequence 3185 BP; 676 A; 883 C; 945 G; 681 T; 0 other;

  Query Match             31.1%;  Score 790.4;  DB 24;  Length 3185;
  Best Local Similarity   92.3%;  Pred. No. 7.8e-181;
  Matches  855;  Conservative    0;  Mismatches   66;  Indels    5;  Gaps    2;

Qy          4 CGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTG 63
              |||  |   |||||||||||| ||||||||||||||||||||||||||||||||||||| |
Db         52 CGCATGAAGCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCG 111

Qy         64 GAGGGCCGATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGC 122
              | | ||  |  || ||| |||||||||||||| |||||||||||||| ||| ||||||||||
Db        112 GGGCGCTGGGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGC 171

Qy        123 CGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 182
              | ||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        172 CCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 231

Qy        183 TTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGG 242
              ||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db        232 TTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGG 291

Qy        243 CCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCG 302
              |||||||||||||||| |||||||| |||||||| |||||||| |||||||||||||| |
Db        292 CCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTG 351

Qy        303 ACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCT 362
              |||| |||||||||||||| || |||||||||||||||||| |||||||||||||| ||||
Db        352 ACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCT 411

Qy        363 GCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCC 422
              |||| ||||| |||||||||||||| |||||||||||||||||||||| |||||||| || |
Db        412 GCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTC 471

Qy        423 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTC 482
              |||||||||||||||||||||||||||||||| ||||||| || ||||||||||||||||
Db        472 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTC 531

Qy        483 GAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACC 542
              |||||||||||||  ||||||| |||||||||||||||||||| ||||| ||||||||||
Db        532 GAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACC 591

Qy        543 TCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTA 602
              |||||||||||||||||||||||||||||||||||| |||||||||| |||||||||| || |
Db        592 TCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGCCCCA 651

Qy        603 ATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCT 662
              |||||||||||||||||||||| |||||||||||||||||||||| ||||| |||||||
Db        652 ATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCT 711

Qy        663 GCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGT 718
              |||||||||||||||||||| |||||||||||||    ||||||||||||||||||||
Db        712 GCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGT 771

Qy        719 GTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778
              ||||||| || |||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db        772 GTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAG 831

Qy        779 GCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCT 838
              |||| |||||||||||||||||||||||||||||||||||||| |||||||| || ||||
Db        832 GCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCT 891

Qy        839 TTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTG 898
              |  |||||||||| |||||||||||||||||| |||||||||||||||||||| |||||||
Db        892 TCATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTG 951

Qy        899 ACATCAAGGCAGTGAAGGCCAACTGG 924
              |||||||||||||||| |||||||||
Db        952 ACATCAAGGCAGTGAAAGCCAACTGG 977


RESULT 5
AAC90014
ID   AAC90014 standard; cDNA; 1126 BP.
XX
AC   AAC90014;
XX
```

DT 09-MAR-2001 (first entry)
XX
DE Clone HNTMH27 coding sequence.
XX
KW Gene therapy; human; 4 transmembrane superfamily receptor protein;
KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW bacterial infection; viral; fungal; ss.
XX
OS Homo sapiens.
XX
PN WO200070076-A1.
XX
PD 23-NOV-2000.
XX
PF 18-MAY-2000; 2000WO-US13504.
XX
PR 19-MAY-1999; 99US-0135122.
PR 03-JUN-1999; 99US-0137797.
PR 11-JUN-1999; 99US-0138573.
PR 18-AUG-1999; 99US-0149447.
PR 28-JAN-2000; 2000US-0178770.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;
PI Rosen CA;
XX
DR WPI; 2001-007502/01.
DR P-PSDB; AAB49504.
XX
PT Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT superfamily receptor protein, useful for diagnosing, treating and/or
PT preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS Claim 1; Pages 269-270; 297pp; English.
XX
CC The present invention relates to isolated nucleic acids and proteins
CC encoding human soluble 4 transmembrane superfamily receptor protein (see
CC AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
CC nucleic acid. The present sequence is useful for preventing, treating or
CC ameliorating a medical condition and in diagnosing (susceptibility to) a
CC pathological condition e.g. endocrine disorders e.g. Addison's disease,
CC (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,
CC cerebrovascular diseases, neural disorders e.g. Alzheimer's and
CC Parkinson's disease, reproductive disorders, skin disorders e.g.
CC psoriasis, renal system disorders e.g. nephritis, (auto)immune system
CC disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.
CC neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections
CC caused by bacteria, viruses and fungi.
XX
SQ Sequence 1126 BP; 191 A; 352 C; 343 G; 240 T; 0 other;

Query Match 31.0%; Score 787.2; DB 22; Length 1126;
Best Local Similarity 92.6%; Pred. No. 3.1e-180;
Matches 850; Conservative 0; Mismatches 63; Indels 5; Gaps 2;

```
Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
Db         49 GCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCGGGGCGCTG 108

Qy         72 ATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAG 130
Db        109 GGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAG 168

Qy        131 CACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 190
Db        169 CACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTC 228

Qy        191 AACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGT 250
Db        229 AACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGT 288

Qy        251 GAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTG 310
Db        289 GAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTG 348

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
Db        349 TGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTGCATTGGG 408

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
Db        409 GCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTC 468

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
Db        469 TTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAG 528

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
Db        529 CTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAAC 588

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
Db        589 CTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGCCCCAATGACTGG 648

Qy        611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTG 670
Db        649 AACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTG 708

Qy        671 CCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTAT 726
Db        709 CCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTAC 768

Qy        727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTG 786
Db        769 GACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTG 828

Qy        787 GGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGC 846
Db        829 GGCCAGTTTGAGAAATGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGC 888

Qy        847 ATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAG 906
Db        889 ATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAAG 948

Qy        907 GCAGTGAAGGCCAACTGG 924
Db        949 GCAGTGAAAGCCAACTGG 966
```

RESULT 6
ABZ11816
ID ABZ11816 standard; cDNA; 1023 BP.
XX
AC ABZ11816;
XX
DT 20-JAN-2003 (first entry)
XX
DE Human polynucleotide SEQ ID NO 698.
XX
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200270539-A2.
XX
PD 12-SEP-2002.

```
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP69599.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders  -
XX
PS   Claim 1; SEQ ID NO 698; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1023 BP; 190 A; 300 C; 293 G; 240 T; 0 other;

  Query Match             30.5%;  Score 773.4;  DB 24;  Length 1023;
  Best Local Similarity   88.1%;  Pred. No. 6.5e-177;
  Matches  922;  Conservative    0;  Mismatches   91;  Indels   34;  Gaps    6;

Qy          89 GGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGCACCAGCACTTCCAGGAA 148
               |||| |||||||||||||| ||| |||||||||||| |||||||||||||| |||||||||
Db           1 GGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCCCGGCAAGCACCAGCATTTCCAGGAA 60

Qy         149 CCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTG 208
               || ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db          61 CCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTCTTCTGGGTG 120

Qy         209 CTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCC 268
               |||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db         121 CTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGCCTGGGGTGAGAAGGGCGTTCTCTCG 180

Qy         269 AACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC-TGTTTGTAGTGGT 327
               |||||||| |||||||| |||||||||||||| ||||| ||||||| ||||||| || ||
Db         181 AACATCTCAGCGCTGACAGATCTGGGAGGCCTTGACCCCGTGTGGCTTGTTTGTGGTAGT 240

Qy         328 TGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGG--CTGCATCGGGGCTCTCCGGGAGAAC 385
               |||||||||||||| |||||||||||||| ||      || ||||| ||||||||||||
Db         241 TGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGGCTGCAATTGGGGCCCTCCGGGAGAAC 300

Qy         386 ACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCA 445
               || |||||||||||||||||||| |||||||| || ||||||||||||||||||||||||
Db         301 ACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCTCATCTTCTTCCTGGAGCTGGCA 360

Qy         446 ACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATT 505
               ||||||||| ||||||| || |||||||||||||||||||||||||||  |||||||
Db         361 ACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCGAGACCAGCTCAACCTCTTCATC 420

Qy         506 AACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCT 565
               |||||||||||||||||||| ||||| |||||||||||||||||||||||||||||||||
Db         421 AACAACAACGTCAAGGCCTACCGGGACGACATTGACCTCCAGAACCTCATTGACTTTGCT 480

Qy         566 CAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTAT 625
               |||||||| |||||||||||||||||||||| || ||||||||||||||||||||||||
Db         481 CAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGGAACCTCAATATCTAC 540

Qy         626 TTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGT 685
               |||||||||||||||||||||||||||| ||||| |||||||||||||||||||||||||
Db         541 TTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTGCCCTTCTCCTGCTGC 600

Qy         686 GTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAA 741
               |||||||||||||    |||||||||||||||||||||||||| || ||||||||||||
Db         601 GTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTACGACGTCCGGCTCAAA 660

Qy         742 CTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAG 801
               |||||||||||||||||||||| ||| |||||||||||||| ||||||||||||||||||
Db         661 CTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTGGGCCAGTTTGAGAAG 720

Qy         802 TGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAG 861
               |||||||||||||||||||||| |||||||| || |||||  |||||||||| |||||||||
Db         721 TGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGCATCGCCCTCCTCCAG 780

Qy         862 ATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCCAAC 921
               |||||||| |||||||||||||||||||| ||||||||||||||||||||||| ||||||
Db         781 ATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAAGGCAGTGAAAGCCAAC 840

Qy         922 TGGATCAAACATGATGATGGCTACAAACTACTCAAATAAACAAAACCTTGAAAACCACTG 981
               |||| ||||    |||||                          || ||||||||||||
Db         841 TGGAGCAAATGGAATGATG----------------------------ACTTTGAAAACCACTG 875

Qy         982 GCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGGGCCTCAGCCGTGCCGTCTGCC 1041
               |||||||||||||| || ||||| |  |   | || |||||||||| |  |  |||| |
Db         876 GCTTACGCCCACCATTTCCGAGGTCCTGTCCACGGCGGGGCCTCAGCAGAACTCTCTGAC 935

Qy        1042 TGGGGCCCC-AGCCCAGACCCACCCTGCCAACATGTTTTC-TTGGCCTGGGTAGTACATA 1099
               |||||||||  |||| | ||||||| ||| |||||||||| |||||||||| ||  |||
Db         936 TGGGGCCCCTGGCCCGGCCCCACCCAGCCGACATGTTTTCTTTGGCCTGGGTGGTTTATA 995

Qy        1100 CGATGAGCCAACCTTTAAAACTTGGCA 1126
               |  |||||||||||||||| |||| |
Db         996 CCCTGAGCCAACCTTTAAAAATTGGTA 1022


RESULT 7
ACC44092
ID   ACC44092 standard; DNA; 1567 BP.
XX
AC   ACC44092;
XX
DT   13-JUN-2003  (first entry)
XX
DE   Rat gamma-hydroxybutyrate receptor cDNA.
XX
KW   gene; rat; gamma-hydroxybutyrate; anti-epileptic; anxiolytic; ds;
KW   antineurodegeneration; antipsychotic; brain; dopamine; opioid; GABA;
KW   gamma-aminobutyric acid; diagnosis; epilepsy; anxiety; sleep disorder;
KW   behavioral disorder; neurodegeneration; Parkinson's disease; psychosis;
KW   schizophrenia.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             22..1560
FT                   /*tag= a
FT                   /product= "gamma-hydroxybutyrate receptor"
XX
PN   WO200078948-A2.
XX
PD   28-DEC-2000.
XX
PF   19-JUN-2000; 2000WO-FR01687.
```

XX
PR 18-JUN-1999; 99FR-0007784.
XX
PA (UYPA-) UNIV PASTEUR LOUIS.
XX
PI Andriamanpandry C, Maitre M;
XX
DR WPI; 2001-091570/10.
DR P-PSDB; ABP98695.
XX
PT New rat brain gamma-hydroxybutyrate receptor and its encoding nucleic
PT acid, useful for identifying agents for treating e.g. epilepsy -
XX
PS Claim 5; Page 64-66; 66pp; French.
XX
CC This sequence represents the cDNA sequence encoding a novel rat
CC gamma-hydroxybutyrate receptor (GHBR) isolated from rat brain hypocampal
CC cell cDNA library. The invention relates to the isolation of this novel
CC sequence, fragments of it and homologues of the sequence except for those
CC homologues that having GenBank accession numbers AAC 17120 (human
CC tetraspan NET-4), AA615405, AA967250 or AI467230. GHBR is the receptor
CC for gamma-hydroxybutyrate in the rat brain and is involved in regulating
CC dopaminergic, opioid and GABA(gamma-aminobutyric acid)ergic activities.
CC The nucleic acid that encodes the protein is used: (1) as primers or
CC probes for detection/amplification, particularly for screening gene
CC libraries to identify promoters and regulators of the GHBR gene; (2) for
CC expression of recombinant polypeptides; and (3) to detect allelic
CC variants, mutations, deletions, loss of heterozygosity or genetic
CC abnormalities in the GHBR gene (for diagnosing diseases, or
CC susceptibility, associated with abnormal expression of GHBR). The
CC protein (or cells and transgenic animals expressing it) is used:
CC (1) to screen for agents that interact with GHBR; (2) to study
CC expression/activity of the receptor, including its interaction with
CC other compounds; and (3) to raise antibodies (Ab) specific for GHBR.
CC The Ab are used to detect/measure (I) in diagnostic immunoassays.
CC The agents are used to prevent or treat diseases associated with abnormal
CC expression or activity of GHBR, particularly those involving cerebral
CC GABA(gamma-aminobutyric acid)ergic and/or dopaminergic activities, e.g.
CC epilepsy, anxiety, sleep or behavioral disorders, withdrawal from
CC addictive drugs, neurodegeneration (Parkinson's disease), psychoses,
CC schizophrenia and regulation of secretion of hormones (growth hormone
CC and prolactin) that are under dopaminergic control.
XX
SQ Sequence 1567 BP; 281 A; 436 C; 480 G; 370 T; 0 other;

Query Match 20.1%; Score 511; DB 23; Length 1567;
Best Local Similarity 85.9%; Pred. No. 2.3e-113;
Matches 602; Conservative 0; Mismatches 95; Indels 4; Gaps 3;

```
Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
Db         28 GCCGCAGCCGCCGGGCTAGGCCCGGGGCGGCTCCACGCCTGGGCAGCCCGCAGAGGGCTG 87

Qy         72 ATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGC 131
Db         88 GGCCGGTTCCCGGCTCGGGTTCCCAGGGCGGCAGGTGGCCGCTCACCATGCCCGGCAAGC 147

Qy        132 ACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 191
Db        148 ATCAGCAATTCCAGGACCCTGAGGTTGGCTGCTGCGGGAAATACTTTCTGTTTGGCTTCA 207

Qy        192 ACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTG 251
Db        208 ACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTG 267

Qy        252 AGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGAT-CTGGGAGGCCTCGACCCTGTG 310
Db        268 AGAAGGGCGTTCTTTCCAACATCTCAGGGCGGACAGATCCAAGGCGGTCTTGACCCCGTG 327

Qy        311 TGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGG 370
Db        328 TGGCTGTTTGTGGTGATTGGGGGAATCATGTCAGTGCTGGGCTTTGCCGGCTGCATTGGG 387

Qy        371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC 430
Db        388 GCCCTCCGGGAAAACACCTTCCTGCTCAAATTTTTCTCTGTGTTCCTCGGCCTCATCTTC 447

Qy        431 TTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAG 490
Db        448 TTCCTGGAGCTGGCGGCCGGGATCCTGGCCTTCGTGTTCAAGGATTGGATCCGAGACCAA 507

Qy        491 CTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAAC 550
Db        508 CTTAACCTCTTCATCAACAACAATGTCAAAGCCTACCGGGACGATATTGACCTTCAGAAC 567

Qy        551 CTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGG 610
Db        568 CTTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGGCCCAATGACTGG 627

Qy        611 AACCTCAATATC--TATTTCAACTGCACTGACTT-GAACCCGAGCCGAGAGCGCTGCGGG 667
Db        628 AACCTCAACATCCGGACTTCAACTGCACTGACTTCAAACCCAAGCCGCGAGCGCTGTGGG 687

Qy        668 GTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTC 708
Db        688 GTGCCCTTCTTCTGCTGGGTAAGGACCCTGCGGAAGACGTC 728
```

RESULT 8
AAH98169
ID AAH98169 standard; cDNA; 1735 BP.
XX
AC AAH98169;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 26.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR P-PSDB; AAM23510.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 200-201; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA

```
CC   of the invention.
XX
SQ   Sequence 1735 BP; 363 A; 463 C; 528 G; 381 T; 0 other;

  Query Match            17.7%;  Score 449.2;  DB 22;  Length 1735;
  Best Local Similarity  64.6%;  Pred. No. 2.1e-98;
  Matches 1106;  Conservative    0;  Mismatches  473;  Indels  133;  Gaps   24;

Qy       732 CCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCA 791
            ||||||||||||||||||||||||||||||||| ||| ||||||||||||| ||||||||
Db         1 CCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTGGGCCA  60

Qy       792 GTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGC 851
            |||||||||||||||||||||||||||||| |||||||| || |||||  ||||||||||
Db        61 GTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGCATCGC 120

Qy       852 TCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGT 911
            ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       121 CCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAAGGCAGT 180

Qy       912 GAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATAAACAAAACCTTG 971
            ||| |||||||||| ||||    ||||||                          || |||
Db       181 GAAAGCCAACTGGAGCAAATGGAATGATG------------------------ACTTTG 215

Qy       972 AAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAGGGCCTCAGCCGT 1031
            |||||||||||||||||||||||| || ||||| |  |   | || |||||||||| |
Db       216 AAAACCACTGGCTTACGCCCACCATTTCCGAGGTCCTGTCCACGGCGGGGCCTCAGCAGA 275

Qy      1032 GCCGTCTGCCTGGGGCCCC-AGCCCAGACCCACCCTGCCAACATGTTTTC-TTGGCCTGG 1089
               |||| ||||||||||  |||| | |||||||| ||| |||||||| |||||||||
Db       276 ACTCTCTGACTGGGGCCCCTGGCCCGGCCCCACCCAGCCGACATGTTTTCTTTGGCCTGG 335

Qy      1090 GTAGTACATACGATGAGCCAACCTTTAAAACTTGGCATATTTCATGTAAAAGTCCAGATC 1149
            || ||  |||| |||||||||||||||||| |||| | |||||| ||||||||||||||
Db       336 GTGGTTTATACCCTGAGCCAACCTTTAAAAATTGGTAGATTTCACATAAAAGTCCAGATC 395

Qy      1150 CCCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGCTTCGTCTCC 1209
            | |||| |||  |||||||| ||| | ||| ||  |||||||||  ||||  |    ||
Db       396 CACAGCTTCTCTTGAAGAATGACCACCTGGCTACGCCGGCTCTTCGGTGGCAACACTACC 455

Qy      1210 TGGGATGTGCGCTTCCTGTTCTCTGAGGGACCCACCCTCACCCGTGTCC---------TG 1260
            |||||       ||  ||    ||  |  |||| |||| |     ||  | | |         ||
Db       456 TGGGACACTGCCTCCCCAGTCACCAAGGGCCCCAGCTGGCCCGTTCTACTCACCTAAGTG 515

Qy      1261 CCTGCCTGACCCTGGAGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTCCCCCTGC 1320
            ||  |    || || |  || ||||||||||||  |||||||||| | ||  |  |||||
Db       516 CCGCCTGACCCTTGTACACTAGGAGCTGGCCTC--CCACCTCTGCAGGGTTATTTCCTGC 573

Qy      1321 AAATGCTGCAAGGCTGCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAGAATTGG 1380
            |  |    | |||| |||| |||||||  |  |  ||  | ||  | ||||| | |||
Db       574 ACCT----CGAGGCCGCTGCGGGCCAATCTGGAGTGAAACACGGGGACCTGAAGGA-TGG 628

Qy      1381 GGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCCTGGGGG 1440
            |||||||| || ||    ||||||   ||||||||||||||    |||| | ||||||
Db       629 AGAGGCTGGACCCCGCTTTGAAGAGGGTGCAGCCTGGGAAGGG--CGGCCTTGCTGGGGA 686

Qy      1441 CCAAGATGGGTGCCACCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATCAGGAAG 1500
            |   | |||| |  |  |||||||||||||| | | |  |||||||||| | ||||||
Db       687 CTGCGGTGGGAG-TAGAGTGCCCAGGAGAGGGTCTGAGGGGTGGGATGGGGGTCAGGACA 745

Qy      1501 GTTTTGGGCAGGACGTAGCTGGAAGCCTGAGCTTGTCACCCATGGGGATGGGGAGAGCCC 1560
            |||||    | || ||||||||||||                ||||  ||  |   ||
Db       746 ATTTTGCAAAAGAAGTAGCTGGAAGCC---------------ATGGGACTGGCGGGAGCC 790

Qy      1561 TGTTTGAGGGCGGCTGATGGTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTTCC 1620
            ||||||  |||     ||||||  |     |         | |||    ||||||  ||||||
Db       791 TGTTTGGGGGATCTGGATGGTTGACTCC--------TAGGAGTCAAGTTCAGCATCTTCG 842

Qy      1621 CAGTGGCCTGTAGAGTTGCCTCCTGACCACTAGAGGGCGCGCCCACACAGCATTACCTGG 1680
            | |||| ||| |||| |||||   ||  |||||||||| ||||  | |  || | |||||
Db       843 CCGTGG-CTGCAGAGCTGCCTGATGGGCACTAGAGGGCACGCCAGCCCCACACTCCCTGG 901

Qy      1681 GTCTGCCTTTCCTAGGACAACCCCACCCAGTACAGCCCTGTGCCTGGTGTGTCCACCCTG 1740
```

```
            ||||| ||| |   |  | | |  | |||| ||||                    ||||
Db       902 GTCTGGCTTCCTCCCGCAACCTCACTCTAGTAGAGCCTG--------------TGCCTG 946

Qy      1741 CTTACTAGTTCTTTGGGTTTCATGGAATTTACAAGCTTCTAAAGGAGCAGAGTGGCTCAG 1800
            | ||||||  || |||| |||   || |||   |  |||| |  ||||  | ||||||||
Db       947 CCTACTAGCGCTCTGGGGTTCGGAGAGTTTGGGAATTTCTCA--GAGCCAACTGGCTCAG 1004

Qy      1801 ATTGGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGTGGAGTATT-- 1858
               | ||||||  ||||| ||||  |||||  || ||   |  | | ||||  | |  |
Db       1005 GCTTGGGAAGGCTGGCTGCTGCCCTCAGCTCCGCCTCATCAGCTATGTGAAGGGGTGTGT 1064

Qy      1859 ------TGTGAATCAAAGGAGAGGTTTGGCCTAGTGCCCAGTCTTTTAACTTAGATGCCC 1912
                  ||   |              ||| || || |  ||   |  ||||  |||||||
Db       1065 ATGGAGTGATCCTGCCGCCCCCTCCCTGGGCTGGTCCAGAGATCTCAAACTCCGATGCCC 1124

Qy      1913 TCAGGGCCGGGTGGGTTATAAAA----ATAAAGTAGGCCTTTGAGCTGTGAGGCCTTTGG 1968
               |||||  ||  ||| |  ||    || ||  ||||| ||||| || ||| |   |
Db       1125 CTGGGGCCACGTATGTTGTGTAAATGGATGAAACAGGCCCTTGAGTTGGGAGCCTGCTTC 1184

Qy      1969 GACTTTAATTTTTCCCACTATTCCTGGAGATGGGACATAGAGAGACATTGCTTTGTGCTG 2028
               || |  |||||||| || |||||||                |   | | ||| ||
Db       1185 ACTTTGA--CTTTCCCACTGTTGCTGGAGACA-------------AAGACATCGTGATG 1228

Qy      2029 AGAAATACTTGCATGATTGAGTCTGAGTCGCTAAGGGCAACTGGCCTTGAGTGACATCAA 2088
               ||| | | ||     | |       ||||| ||| || |||  |||||||    |  |
Db       1229 AGAGAAAGTTCGCACAAT-----CTAGTCGGTAACAGCCACTTTCCTTGAGACCAAGAGA 1283

Qy      2089 GGGGTGGTGGGGACTGTGGCAAACCACAGATTCCCACCTGAAATTGGTGGCTGTC--CTT 2146
               | |  | |||    | || | | ||| | | |||  |||| | |||  |||| |   ||
Db       1284 GTGCGGTGGGGATGGGGGGGAGAGCACGGGTCCCCGTCTGACAGTGGCCGCTGCCATATT 1343

Qy      2147 CCGTTGGGGCTAATGGCTGTACAGCGAGAATGTAGGTAGGTCTGTCTAATGGGAGAAGTC 2206
               | | ||  |||||| ||| |    | | |||| ||    | ||||||  |||| ||
Db       1344 CAGGTGTAGCTAATTGCTCTGGTGTGGGAATGCAG--------GCCTAATGACAGAAATC 1395

Qy      2207 TGGAGAAGCCAAGAAGCTAGATTTTTCATGTGAACTATCCCGAGTTTTAAGTTGTTTGCA 2266
               ||||||||||  ||  |||||| | ||||||  | ||| || |||| |   |  |||
Db       1396 TGGAGAAGCCAGAAATACAGATTTGT-ATGTGAGATGTCCTGATTTTTTAAGTTGTTGGC 1454

Qy      2267 GCTAATGAGAAAAACCTCTTAAACCCTGATAGTCAAAAGGTGTGGGGGCCATCTTTGACA 2326
               |||  |    |       || |          | ||||||  ||  ||| ||||| |
Db       1455 AGAAATTAATTCAGAAATCAAATCTGCAGGCCAAACAAGGTGCAGGACCCAGCTTTGGCC 1514

Qy      2327 CCTCCCCCACCATAGGTCCCTCAGGGACAGTGCCCCATGGGAGCCCTGGTGAGTCCACGG 2386
               ||   ||| |  |||||||||| ||||||||  ||  |||| | |||||     ||
Db       1515 CCATGCCC-CTGTAGGTCCCTCTGGGACAGTCACCGCTGGG-GTCCTGGCTGCTCTGTCA 1572

Qy      2387 TTCAGGAATGCTGGGAACTGCTGCAGGTGGGC 2418
               || ||| |||||||| ||||||||  ||  |||
Db       1573 TTGAGGGATGCTGGGCACTGCTGCCGGGTGGC 1604
```

```
RESULT 9
AAC98195
ID   AAC98195 standard; cDNA; 1655 BP.
XX
AC   AAC98195;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen nucleotide sequence SEQ ID NO:205.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
```

```
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; AAB53438.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 628-629; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1655 BP; 371 A; 415 C; 437 G; 425 T; 7 other;

  Query Match             17.4%;  Score 441.6;  DB 21;  Length 1655;
  Best Local Similarity   73.2%;  Pred. No. 1.4e-96;
  Matches  581;  Conservative    0;  Mismatches  209;  Indels    4;  Gaps    1;

Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   |||||||||| | ||||||||||
Db        287 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 346

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||      || || |  || || || ||||| |||  ||| |
Db        347 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 406

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | || ||||| |||||||||||| |  | |||||||| || ||| | ||||| || ||||
Db        407 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 466

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | |  |||| ||||| || ||||| || || |
Db        467 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGCGC 526

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCC 434
              | ||||| |||||||||| |||||||||| || |||||||| ||  | || |||||||
Db        527 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATTATTTTCTTCC 586

Qy        435 TGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCA 494
              |||||||  |  | ||  |  | || || || ||||| ||||||||   |||||||||
Db        587 TGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTGT 646

Qy        495 ATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCA 554
              ||||||| || ||||||||| |||  || |||||||||||||||||  | || |||||||
Db        647 ATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTCA 706

Qy        555 TTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACC 614
              | |||||  | ||||||||||||   ||||| || ||    ||  || |||| |||||||

Db        707 TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAACC 766

Qy        615 TCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCT 674
              | ||||| || ||||| ||||| || |  ||  | || |||||||| || || || || |
Db        767 TAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCAT 826

Qy        675 TCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATG 730
              ||||||||||     |  || || ||    |||||| ||||||| ||||||||||||||||
Db        827 TCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGATG 886

Qy        731 TCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCC 790
              | |||  ||||  ||  | || ||||||      ||  |||| ||||||||||||  ||
Db        887 CCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCCC 946

Qy        791 AGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCG 850
              |||||||||||||| ||||||||||. | | | | || |||||  | ||  | ||||| |
Db        947 AGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATTG 1006

Qy        851 CTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAG 910
              |  | || ||||| ||||| || ||||||||||||||  | || || || ||| | || |
Db       1007 CATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCTG 1066

Qy        911 TGAAGGCCAACTGG 924
              | | ||| | ||||
Db       1067 TCAGGGCGAGCTGG 1080


RESULT 10
AAC90015
ID   AAC90015 standard; cDNA; 1174 BP.
XX
AC   AAC90015;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Clone HE8EJ16 coding sequence #1.
XX
KW   Gene therapy; human; 4 transmembrane superfamily receptor protein;
KW   endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW   reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW   bacterial infection; viral; fungal; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070076-A1.
XX
PD   23-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13504.
XX
PR   19-MAY-1999;   99US-0135122.
PR   03-JUN-1999;   99US-0137797.
PR   11-JUN-1999;   99US-0138573.
PR   18-AUG-1999;   99US-0149447.
PR   28-JAN-2000; 2000US-0178770.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Fan P,  Roschke V,  Shi Y,  Komatsoulis GA;
PI   Rosen CA;
XX
DR   WPI; 2001-007502/01.
DR   P-PSDB; AAB49505.
XX
PT   Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT   superfamily receptor protein, useful for diagnosing, treating and/or
PT   preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS   Claim 1; Page 270; 297pp; English.
XX
CC   The present invention relates to isolated nucleic acids and proteins
CC   encoding human soluble 4 transmembrane superfamily receptor protein (see
CC   AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
```

```
CC   nucleic acid. The present sequence is useful for preventing, treating or
CC   ameliorating a medical condition and in diagnosing (susceptibility to) a
CC   pathological condition e.g. endocrine disorders e.g. Addison's disease,
CC   (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,
CC   cerebrovascular diseases, neural disorders e.g. Alzheimer's and
CC   Parkinson's disease, reproductive disorders, skin disorders e.g.
CC   psoriasis, renal system disorders e.g. nephritis, (auto)immune system
CC   disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.
CC   neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections
CC   caused by bacteria, viruses and fungi.
XX
SQ   Sequence 1174 BP; 306 A; 259 C; 291 G; 316 T; 2 other;

  Query Match             16.5%;  Score 418.6;  DB 22;  Length 1174;
  Best Local Similarity   73.0%;  Pred. No. 4.6e-91;
  Matches  580;  Conservative    0;  Mismatches  209;  Indels    6;  Gaps    3;

Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        103 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 162

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||      || || |  || || || ||||| |||  ||| |
Db        163 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 222

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | || ||||| |||||||||||| |  | |||||||| || ||| | ||||| || ||||
Db        223 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 282

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | |  |||| ||||| || ||||| || ||
Db        283 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGC-G 341

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTC-TTC 433
              | ||||| |||||||||||| |||||||||||  || |||||||  ||  | ||  ||| |||
Db        342 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATTATTTTCTTTC 401

Qy        434 CTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTC 493
              ||||||||  |  | ||  |  | || || || ||||| ||||||||   |||||||||
Db        402 CTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTG 461

Qy        494 AATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTC 553
              |||||||  || |||||||||  |||   || ||||||||||||||||  | || ||||||
Db        462 TATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTC 521

Qy        554 ATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAAC 613
              || |||||  | ||||||||||||   ||||| || ||    ||  || |||| ||||||
Db        522 ATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAAC 581

Qy        614 CTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCC 673
              || ||||| || ||||| ||||| || |  ||  | || |||||||| || || || ||
Db        582 CTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCA 641

Qy        674 TTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGAT 729
              ||||||||||     |  || || ||    |||||| ||||||| ||||||||||||||
Db        642 TTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGAT 701

Qy        730 GTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGC 789
              | | |||  ||||  ||  | || ||||||      ||  |||| ||||||||||||  |
Db        702 GCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCC 761

Qy        790 CAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATC 849
              ||||||||||||||| ||||||||||  | | | | || |||||  | ||  | |||||
Db        762 CAGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATT 821

Qy        850 GCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCA 909
              ||  | || ||||| ||||| || ||||||||||||||  | || || ||| | ||
Db        822 GCATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCT 881

Qy        910 GTGAAGGCCAACTGG 924
              || | ||| | ||||
Db        882 GTCAGGGCGAGCTGG 896
```

```
RESULT 11
AAC90020
ID   AAC90020 standard; cDNA; 1178 BP.
XX
AC   AAC90020;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Clone HE8EL16 coding sequence #2.
XX
KW   Gene therapy; human; 4 transmembrane superfamily receptor protein;
KW   endocrine; cardiovascular; cerebrovascular disease; neural disorder;
KW   reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;
KW   bacterial infection; viral; fungal; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070076-A1.
XX
PD   23-NOV-2000.
XX
PF   18-MAY-2000; 2000WO-US13504.
XX
PR   19-MAY-1999;   99US-0135122.
PR   03-JUN-1999;   99US-0137797.
PR   11-JUN-1999;   99US-0138573.
PR   18-AUG-1999;   99US-0149447.
PR   28-JAN-2000; 2000US-0178770.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Fan P,  Roschke V,  Shi Y,  Komatsoulis GA;
PI   Rosen CA;
XX
DR   WPI; 2001-007502/01.
DR   P-PSDB; AAB49510.
XX
PT   Isolated nucleic acid molecule encoding human soluble 4 transmembrane
PT   superfamily receptor protein, useful for diagnosing, treating and/or
PT   preventing disorders e.g. Alzheimer's, cancer and arrhythmia -
XX
PS   Claim 1; Pages 274-275; 297pp; English.
XX
CC   The present invention relates to isolated nucleic acids and proteins
CC   encoding human soluble 4 transmembrane superfamily receptor protein (see
CC   AAC90012-C90023 and AAB49502-B49513). The present sequence is one such
CC   nucleic acid. The present sequence is useful for preventing, treating or
CC   ameliorating a medical condition and in diagnosing (susceptibility to) a
CC   pathological condition e.g. endocrine disorders e.g. Addison's disease,
CC   (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,
CC   cerebrovascular diseases, neural disorders e.g. Alzheimer's and
CC   Parkinson's disease, reproductive disorders, skin disorders e.g.
CC   psoriasis, renal system disorders e.g. nephritis, (auto)immune system
CC   disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.
CC   neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections
CC   caused by bacteria, viruses and fungi.
XX
SQ   Sequence 1178 BP; 310 A; 261 C; 291 G; 316 T; 0 other;

  Query Match             15.8%;  Score 402.2;  DB 22;  Length 1178;
  Best Local Similarity   71.6%;  Pred. No. 4.2e-87;
  Matches  571;  Conservative    0;  Mismatches  218;  Indels    8;  Gaps    3;

Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        101 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 160

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||      || || |  || || || ||||| |||  ||| |
Db        161 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 220

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
```

```
Db        221 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 280

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374

Db        281 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGC-G 339

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTC---AGTGTTCCTTGGCCTCATCTTCT 431

Db        340 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTCGTGTTCCTCGGGAATTATTTTCTT 399

Qy        432 TCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGC 491

Db        400 TCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGC 459

Qy        492 TCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACC 551

Db        460 TGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACC 519

Qy        552 TCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGA 611

Db        520 TCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGA 579

Qy        612 ACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGC 671

Db        580 ACCTAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTC 639

Qy        672 CCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATG 727

Db        640 CATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATG 699

Qy        728 ATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGG 787

Db        700 ATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGC 759

Qy        788 GCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCA 847

Db        760 CCCAGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCA 819

Qy        848 TCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGG 907

Db        820 TTGCATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAG 879

Qy        908 CAGTGAAGGCCAACTGG 924

Db        880 CTGTCAGGGCGAGCTGG 896
```

RESULT 12
ABK35735
ID ABK35735 standard; cDNA; 1988 BP.
XX
AC ABK35735;
XX
DT 08-MAY-2002 (first entry)
XX
DE cDNA sequence #126 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; allergic condition; neurodegenerative disorder;
KW liver fibrosis; coagulation disorder; gene therapy; antimicrobial;
KW tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200177289-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10232.
XX
PR 06-APR-2000; 2000US-195605P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;
PI Clark HF, Fechtel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;
XX
DR WPI; 2002-179322/23.
XX
PT Six hundred and twenty three polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT -
XX
PS Claim 1; Page 146; 393pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides
CC a method for producing proteins from these polynucleotide sequences.
CC The proteins are useful for identifying compounds that modulate their
CC activity and production. The sequences of the invention are
CC useful for treating diseases such as hyperproliferative disorders
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),
CC allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.
CC Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.
CC haemophilia), and tumours. The polynucleotide sequences of the
CC invention are also useful in gene therapy. ABK35610-ABK36232 represent
CC the cDNA sequences of the invention that encode for novel human
CC secreted proteins.
XX
SQ Sequence 1988 BP; 503 A; 434 C; 446 G; 604 T; 1 other;

Query Match 14.6%; Score 370; DB 24; Length 1988;
Best Local Similarity 72.4%; Pred. No. 3.2e-79;
Matches 495; Conservative 0; Mismatches 185; Indels 4; Gaps 1;

```
Qy        245 TGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGAC 304

Db          2 TGGAATGAAAAAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGAC 61

Qy        305 CCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGC 364

Db         62 CCAGTTTGGCTCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGC 121

Qy        365 ATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTC 424

Db        122 ATTGGAGCGCTACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATT 181

Qy        425 ATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGA 484

Db        182 ATTTTCTTCCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAA 241

Qy        485 GACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTC 544

Db        242 GACCAGCTGTATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTG 301

Qy        545 CAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAAT 604

Db        302 CAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGAT 361

Qy        605 GACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGC 664

Db        362 GATTGGAACCTAAATATTTACTTCAATTGTCCCGCAGAAGATGTCATCAACACTCAGTGT 421

Qy        665 GGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGT 720

Db        422 GGCGTTCCATTCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGT 481

Qy        721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGC 780
```

```
Db        482 GGCTATGATGCCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGC 541

Qy        781 TGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTT 840

Db        542 TGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTC 601

Qy        841 GTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGAC 900

Db        602 ATAGGCATTGCATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGAT 661

Qy        901 ATCAAGGCAGTGAAGGCCAACTGG 924

Db        662 ATCGAAGCTGTCAGGGCGAGCTGG 685


RESULT 13
AAZ65259
ID   AAZ65259 standard; DNA; 2672 BP.
XX
AC   AAZ65259;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human secreted protein gene 10.
XX
KW   Human; secreted protein; cancer; tumour; developmental abnormality;
KW   foetal deficiency; blood disorder; immune system disorder; inflammation;
KW   autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW   schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW   atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW   digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
KW   therapy; chromosome 10; ds.
XX
OS   Homo sapiens.
XX
PN   WO9958660-A1.
XX
PD   18-NOV-1999.
XX
PF   06-MAY-1999;   99WO-US09847.
XX
PR   12-MAY-1998;   98US-0085093.
PR   12-MAY-1998;   98US-0085094.
PR   12-MAY-1998;   98US-0085105.
PR   12-MAY-1998;   98US-0085180.
PR   18-MAY-1998;   98US-0085906.
PR   18-MAY-1998;   98US-0085920.
PR   18-MAY-1998;   98US-0085921.
PR   18-MAY-1998;   98US-0085922.
PR   18-MAY-1998;   98US-0085923.
PR   18-MAY-1998;   98US-0085924.
PR   18-MAY-1998;   98US-0085928.
PR   18-MAY-1998;   98US-0085925.
PR   18-MAY-1998;   98US-0085927.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Florence K,  Ni J,  Rosen CA,  Carter KC,  Moore PA;
PI   Olsen HS,  Shi Y,  Young PE,  Wei F,  Brewer LA,  Soppet DR;
PI   Lafleur DW,  Endress GA,  Ebner R;
XX
DR   WPI; 2000-062296/05.
DR   P-PSDB; AAY76133.
XX
PT   New isolated human genes and the secreted polypeptides they encode,
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders -
XX
PS   Claim 1; Page 301-302; 475pp; English.
XX
CC   AAZ65250 to AAZ65350 represent 97 isolated human secreted protein genes.
CC   This sequence was found to be present on human chromosome 10.
CC   AAY76124 to AAY76223 represent the secreted proteins encoded by the 97
CC   human genes. The genes and their corresponding secreted polypeptides are
CC   useful for preventing, treating or ameliorating medical conditions,
CC   e.g. by protein or gene therapy. Also pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new genes. Specific
CC   uses are described for each of the 97 genes, based on which tissues they
CC   are most highly expressed in, and include developing products for the
CC   diagnosis or treatment of cancer, tumours, developmental abnormalities
CC   and foetal deficiencies, blood disorders, diseases of the immune system,
CC   autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC   disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC   disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC   disorders, digestive/endocrine disorders, infections and AIDS. The
CC   polypeptides are also useful for identifying their binding partners.
CC   The sequences shown in AAY76224 to AAY76424 represent fragments of the
CC   secreted proteins.
XX
SQ   Sequence 2672 BP; 548 A; 747 C; 708 G; 661 T; 8 other;

  Query Match           14.1%;  Score 357.4;  DB 21;  Length 2672;
  Best Local Similarity  64.3%;  Pred. No. 4e-76;
  Matches  564;  Conservative    3;  Mismatches  304;  Indels    6;  Gaps    2;

Qy         48 CCCAGGGCGGCCCGTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGC 107

Db         99 CCCGGGGCTGCCAGGAATTTCCGAGCCGGGGCCGCGCCGCCGCTGCCCGCCGCCGCGSGC 158

Qy        108 GGCTGCT--CACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTG 165

Db        159 GGATTYTGCTTCTCAGAAGATGCACTATTATAGATACTCTAACGCCAAGGTCAGCTGCTG 218

Qy        166 CGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCT 225

Db        219 GTACAAGTACCTCCTTTTCAGCTACAACATCATCTTCTGRTTGGCTGGAGTTGTCTTCCT 278

Qy        226 GGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGAC 285

Db        279 TGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCGACCTCACCAAAGTGAC 338

Qy        286 CGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGT 345

Db        339 CCGGATGCATGGAATCGACCCTGTGGTGCTGGTCCTGATGGTGGGCGTGGTGATGTTCAC 398

Qy        346 GCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTT 405

Db        399 CCTGGGGTTCGCCGGCTGCGTGGGGGCTCTGCGGGAGAATATCTGCTTGCTCAACTTTTT 458

Qy        406 CTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGT 465

Db        459 CTGTGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGGCCGTGCTGGCCTTCCT 518

Qy        466 ATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTA 525

Db        519 GTTCCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGAGCAACATCAAGTCCTA 578

Qy        526 TCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTG 585

Db        579 CCGGGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGAAAGCTAACCAGTGCTG 638

Qy        586 CGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAA 645

Db        639 TGGCGCATATGGCCCTGAAGACTGGGACCTCAACGTCTACTTCAATTGCAGCGGTGCCAG 698

Qy        646 CCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCG----A 701

Db        699 CTACAGCCGAGAGAAGTGCGGGGTCCCCTTCTCCTGCTGCGTGCCAGATCCTGCGCAAAA 758

Qy        702 TGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGG 761

Db        759 AGTTGTGAACACACAGTGTGGATATGATGTCAGGATTCAGCTGAAGAGCAAGTGGGATGA 818

Qy        762 CTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGAT 821
```

```
                 |||||   ||| ||||||||  |      | | || |  ||||| | |   ||| |
Db        819 GTCCATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGCTCCCGCGGAACATTTA 878

Qy        822 CGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGC 881
              | | |||||||| |||||  | | |||| | ||  | ||||| ||||| |||| ||||||
Db        879 CATTGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATATTTGGCATCTTCCTGGC 938

Qy        882 CCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCC 918
                ||  ||  |   |||||| ||||||||||||||
Db        939 AAGGACGCTGATCTCAGACATCGAGGCAGTGAAGGCC 975
```

RESULT 14
ABX70953
ID ABX70953 standard; cDNA; 2465 BP.
XX
AC ABX70953;
XX
DT 05-MAR-2003 (first entry)
XX
DE Novel human cDNA sequence #178.
XX
KW Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW coagulation disorder; cancer; tumour; inflammatory disease;
KW septic shock; Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW haemostatic; antiinflammatory; expressed sequence tag; EST.
XX
OS Homo sapiens.
XX
PN WO200281731-A2.
XX
PD 17-OCT-2002.
XX
PF 29-JAN-2002; 2002WO-US01222.
XX
PR 30-JAN-2001; 2001US-0774528.
XX
PA (HYSE-) HYSEQ INC.
PA (GOOD/) GOODRICH R W.
XX
PI Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
XX
DR WPI; 2003-058563/05.
XX
PT Novel polypeptide useful for treating neurodegenerative diseases,
PT myeloid or lymphoid cell disorders, bone disorders, mechanical and
PT traumatic disorders, coagulation disorders, and inflammatory diseases
PT -
XX
PS Claim 1; Page -; 612pp; English.
XX
CC This invention relates to the cDNA sequences encoding an isolated
CC novel human polypeptide. The protein encoded by the nucleic acid of
CC the invention is useful for treating central and peripheral nervous
CC system diseases (e.g. peripheral neuropathy, Huntington's disease,
CC amyotrophic lateral sclerosis); neurodegenerative diseases (e.g.
CC Parkinson's disease, Alzheimer's disease); autoimmune disease (e.g.
CC systemic lupus erythematosus, rheumatoid arthritis, insulin-dependent
CC diabetes mellitus); myeloid or lymphoid cell disorders (e.g. anaemia
CC and thrombocytopaenia); wounds, ulcers, burns; bone disorders (e.g.
CC osteoporosis, osteoarthritis); mechanical and traumatic disorders (e.g.
CC stroke, head trauma); lung or liver fibrosis; reperfusion injury in
CC various tissues; bacterial, viral or fungal infections; allergic
CC conditions such as allergic rhinitis, asthma; coagulation disorders
CC (e.g. haemophilia); cancer and tumours; and inflammatory diseases (e.g.
CC septic shock, Crohn's disease, anaphylaxis). The protein may be used to
CC inhibit the growth, infection or function of infectious agents such as
CC bacteria, fungi, viruses, or to effect bodily characteristics,
CC biorhythms or circadian cycles of rhythms. The protein may also
CC have proliferation/differentiation, stem cell growth factor,
CC haematopoiesis regulation, immune stimulating or suppressing,
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC and antiinflammatory activities. The cDNA sequences of the invention are
CC useful for expressing recombinant protein for analysis. The present
CC sequence represents a novel human cDNA sequence of the invention,
CC this sequence is an expressed sequence tag (EST) and was identified
CC using subtractive hybridisation.
XX
SQ Sequence 2465 BP; 474 A; 718 C; 673 G; 600 T; 0 other;

Query Match 14.0%; Score 354.8; DB 25; Length 2465;
Best Local Similarity 64.5%; Pred. No. 1.7e-75;
Matches 564; Conservative 0; Mismatches 302; Indels 8; Gaps 2;

```
Qy         49 CCAGGGCGGCCCGTGGAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCG 108
              || ||| | | |   | ||||  ||  ||| | ||| | |   ||| ||  |    |
Db         43 CCCGGGTCGACGATTTCGTGCCGGGCCGCGCCGCCGCTGCCGCCGCCGCGCGCGGATTCT 102

Qy        109 GCTGCTCACCATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGG 168
              ||| ||||   | |   | | |   | |     | |  || ||||| |||||||
Db        103 GCTTCTCA----GAAGATGCACTATTATAGATACTCTAACGCCAAGGTCAGCTGCTGGTA 158

Qy        169 GAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGC 228
              || ||| |||| ||  ||| ||||||  | ||||||  ||   ||||   | ||||| |
Db        159 CAAGTACCTCCTTTTCAGCTACAACATCATCTTCTGGTTGGCTGGAGTTGTCTTCCTTGG 218

Qy        229 CATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGA 288
              ||||  || ||||| ||| | || ||||||||  || ||| || || |     |||||
Db        219 AGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCGACCTCACCAAAGTGACCCG 278

Qy        289 TCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCT 348
              ||   ||  |||||||||||  |||| |  |  |||| || |  || |||| |   ||
Db        279 GATGCATGGAATCGACCCTGTGGTGCTGGTCCTGATGGTGGGCGTGGTGATGTTCACCCT 338

Qy        349 GGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTC 408
              ||| || |||||||| | |||||||| |||||||| |  | | |||||| |||||||
Db        339 GGGGTTCGCCGGCTGCGTGGGGGCTCTGCGGGAGAATATCTGCTTGCTCAACTTTTTCTG 398

Qy        409 AGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATT 468
              |    | | |||||||||||||||||||||||    |  | |||||||| | ||
Db        399 TGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGGCCGTGCTGGCCTTCCTGTT 458

Qy        469 CAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCG 528
              | |||||||| |  | |||| | ||       ||| |  | | |||| |||| |||| ||
Db        459 CCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGAGCAACATCAAGTCCTACCG 518

Qy        529 GGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGG 588
              ||| || || || || || |||||||| ||||    |||| ||  |      ||||| ||
Db        519 GGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGAAAGCTAACCAGTGCTGTGG 578

Qy        589 AGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCC 648
              ||     || ||| | |||||| |||||||  |||| |||| ||||  |     | |
Db        579 CGCATATGGCCCTGAAGACTGGGACCTCAACGTCTACTTCAATTGCAGCGGTGCCAGCTA 638

Qy        649 GAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCG----ATGT 704
              |||||||||   |||||||| |||||||||||||| ||    || ||||||    | ||
Db        639 CAGCCGAGAGAAGTGCGGGGTCCCCTTCTCCTGCTGCGTGCCAGATCCTGCGCAAAAAGT 698

Qy        705 CCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTC 764
              | ||||| |||||||| ||||||||| || |  | ||| ||    ||  | | |  ||
Db        699 TGTGAACACACAGTGTGGATATGATGTCAGGATTCAGCTGAAGAGCAAGTGGGATGAGTC 758

Qy        765 CATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGT 824
              |||   ||| |||||||| |       | | || |  ||||| | |   ||| |   | |
Db        759 CATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGCTCCCGCGGAACATTTACAT 818

Qy        825 GGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCA 884
              |||||||| |||||  | | |||| | ||  | ||||| ||||| |||| ||||||
```

```
Db        819 TGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATATTTGGCATCTTCCTGGCAAG 878

Qy        885 GAACCTTGTGAGTGACATCAAGGCAGTGAAGGCC 918
              ||  ||  |    |||||| ||||||||||||||
Db        879 GACGCTGATCTCAGACATCGAGGCAGTGAAGGCC 912


RESULT 15
AAF90629
ID   AAF90629 standard; cDNA; 813 BP.
XX
AC   AAF90629;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Human TANGO 339 cDNA ORF, SEQ ID NO:2.
XX
KW   Secreted protein; transmembrane protein; TANGO; human; drug screening;
KW   activity modulator; expression modulator; cancer; immunological disorder;
KW   cytostatic; immunomodulatory; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200109162-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-US20935.
XX
PR   30-JUL-1999;   99US-0365164.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Fraser CC,  Sharp JD,  Kirst SJ,  Barnes TM,  Wrighton N,  Myers PS;
PI   Pan Y;
XX
DR   WPI; 2001-138647/14.
DR   P-PSDB; AAB87034, AAB87035, AAB87036.
XX
PT   Nucleic acids encoding secreted TANGO and MANGO polypeptides, useful
PT   for the prevention, diagnosis and treatment of, e.g. cancers and immune
PT   disorders -
XX
PS   Claim 2; Page 220; 332pp; English.
XX
CC   The invention relates to novel secreted/transmembrane proteins, and
CC   nucleic acids encoding them. The novel proteins are designated TANGO 339,
CC   TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO
CC   393, TANGO 402, MANGO 346 and MANGO 349 and are of human origin, and a
CC   murine TANGO 393 is also included within the scope of the invention. The
CC   invention also encompasses fragments and variants of the proteins of the
CC   invention, and nucleic acids encoding them. The invention additionally
CC   relates to host cells comprising a nucleic acid of the invention; methods
CC   for the production of a protein of the invention; an antibody specific
CC   for a protein of the invention; methods for detecting a protein or
CC   nucleic acid of the invention; and methods of identifying agents which
CC   bind to or modulate the activity of a protein of the invention. The novel
CC   secreted proteins, nucleic acids encoding them, and antibodies against
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate expression or activity of the secreted
CC   proteins. The secreted proteins of the invention may also be used to
CC   identify modulators of expression or activity, which may be useful in
CC   the treatment of disorders associated with the proteins of the
CC   invention e.g., cancers and immunological disorders. The present
CC   sequence represents cDNA encoding human TANGO protein of the invention.
XX
SQ   Sequence 813 BP; 163 A; 213 C; 235 G; 202 T; 0 other;

  Query Match             13.9%;  Score 353.4;  DB 22;  Length 813;
  Best Local Similarity   67.0%;  Pred. No. 2.3e-75;
  Matches  518;  Conservative    0;  Mismatches  251;  Indels    4;  Gaps    1;

Qy        150 CCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGC 209
              || ||||| |||||||    || ||| |||| ||  ||| ||||||  | |||||| ||
Db         26 CCAAGGTCAGCTGCTGGTACAAGTACCTCCTTTTCAGCTACAACATCATCTTCTGGTTGG 85

Qy        210 TGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCA 269
              ||||   | |||||  |   |||| || ||||| ||| | || |||||||| || |||
Db         86 CTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCG 145

Qy        270 ACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTG 329
              || || |     |||||    ||   ||  ||||||||||  |||| |  |  |||| |
Db        146 ACCTCACCAAAGTGACCCGGATGCATGGAATCGACCCTGTGGTGCTGGTCCTGATGGTGG 205

Qy        330 GAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTT 389
              | |  || |||| |   ||||| || |||||||| | |||||||| |||||||| | |
Db        206 GCGTGGTGATGTTCACCCTGGGGTTCGCCGGCTGCGTGGGGGCTCTGCGGGAGAATATCT 265

Qy        390 TCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAG 449
              | ||||||| |||||||  |    | | |  ||||||||||||||||||||||     |
Db        266 GCTTGCTCAACTTTTTCTGTGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGG 325

Qy        450 GGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACA 509
              |  |||||||| | ||| |||||||| |  | |||| | ||      ||| |  | |
Db        326 CCGTGCTGGCCTTCCTGTTCCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGA 385

Qy        510 ACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGG 569
              |||| ||||| |||| ||||| || || || || || |||||||| ||||    ||||
Db        386 GCAACATCAAGTCCTACCGGGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGA 445

Qy        570 AATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCA 629
              ||  |      ||||| || ||     || ||| | |||||| |||||||  |||| ||||
Db        446 AAGCTAACCAGTGCTGTGGCGCATATGGCCCTGAAGACTGGGACCTCAACGTCTACTTCA 505

Qy        630 ACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCA 689
              | ||||  |     | |  |||||||||   |||||||| |||||||||||||| ||
Db        506 ATTGCAGCGGTGCCAGCTACAGCCGAGAGAAGTGCGGGGTCCCCTTCTCCTGCTGCGTGC 565

Qy        690 GGGACCCTGCG----ATGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGG 745
              || ||||||    | ||  | ||||| |||||||| |||||||| || |  | |||
Db        566 CAGATCCTGCGCAAAAAGTTGTGAACACACAGTGTGGATATGATGTCAGGATTCAGCTGA 625

Qy        746 AGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGC 805
              ||    ||  | | |  ||||   ||| ||||||||  |      | | || |  ||||
Db        626 AGAGCAAGTGGGATGAGTCCATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGC 685

Qy        806 TGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCT 865
              | | |   ||| |   | | |||||||| |||||  | | |||| | ||  | |||||
Db        686 TCCCGCGGAACATTTACATTGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATAT 745

Qy        866 TTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCC 918
              |||| |||| ||||||   ||  ||  |    |||||| ||||||||||||||
Db        746 TTGGCATCTTCCTGGCAAGGACGCTGATCTCAGACATCGAGGCAGTGAAGGCC 798


Search completed: November 21, 2003, 17:31:07
Job time : 658 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 17:13:09 ; Search time 9165 Seconds
(without alignments)
11328.820 Million cell updates/sec

Title: US-09-972-970-2
Perfect score: 2538
Sequence: 1 ccacgcgtccggccgcagcc.....aaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

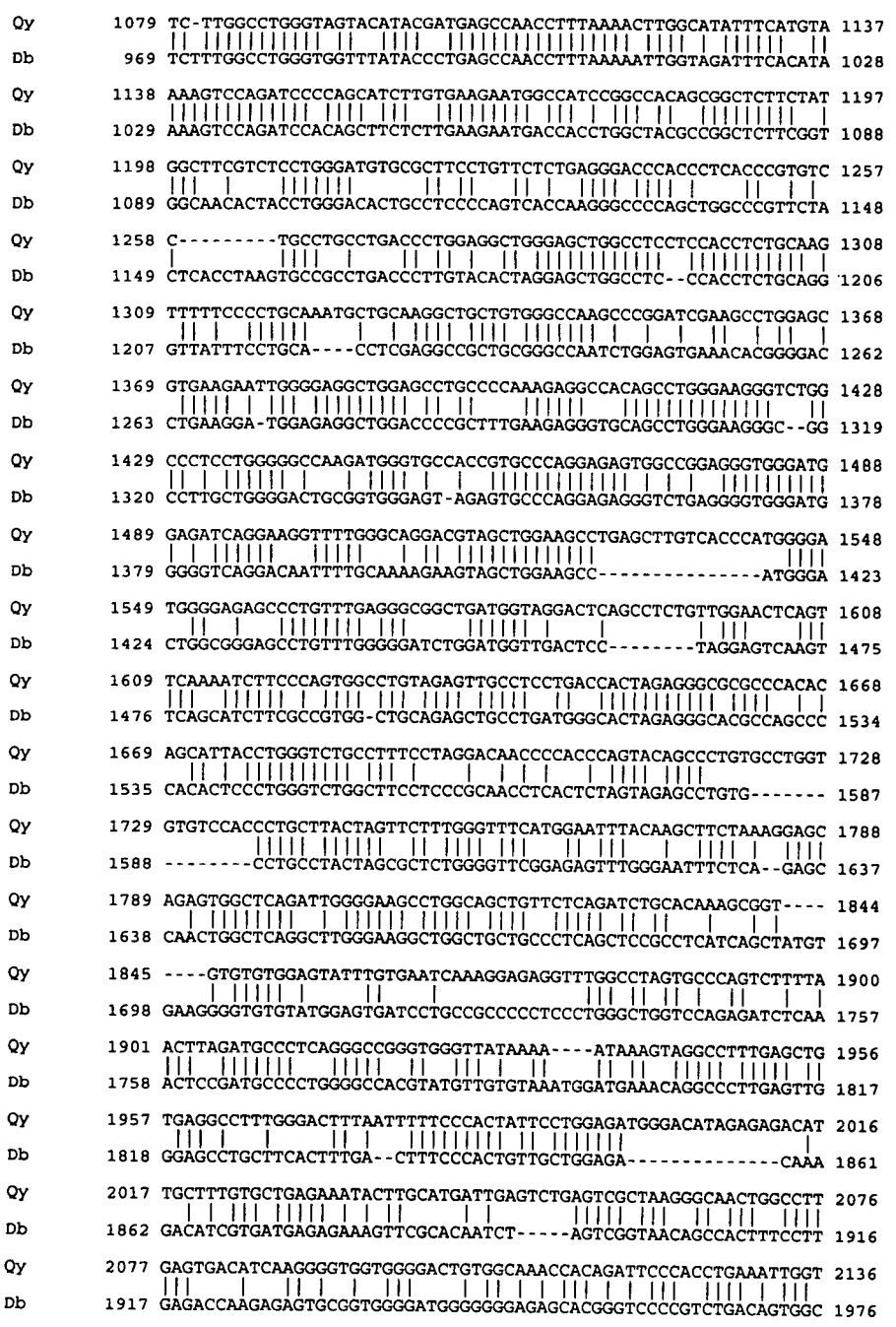
Result No.	Score	% Query Match	Length	DB	ID	Description
1	960.8	37.9	2426	9	BC010405	BC010405 Homo sapi
2	790.4	31.1	3184	6	AX420466	AX420466 Sequence
3	745.4	29.4	1516	10	BC010346	BC010346 Mus muscu
4	710.4	28.0	813	6	AX420468	AX420468 Sequence
5	580.2	22.9	4445	9	AK024427	AK024427 Homo sapi
6	511	20.1	1567	6	AX061778	AX061778 Sequence
7	465.2	18.3	1685	5	BC041304	BC041304 Xenopus l
8	456.4	18.0	3175	10	AF121344	AF121344 Mus muscu
9	441.6	17.4	1405	9	AF065389	AF065389 Homo sapi
10	441.6	17.4	1416	9	BC009704	BC009704 Homo sapi
11	409.2	16.1	1408	9	AF053455	AF053455 Homo sapi
12	368.8	14.5	2428	10	BC025568	BC025568 Mus muscu
13	368.8	14.5	2498	10	BC024611	BC024611 Mus muscu
14	368.8	14.5	2500	10	BC026574	BC026574 Mus muscu
15	353.4	13.9	2553	9	AF311903	AF311903 Homo sapi
16	352	13.9	2556	9	HSM801611	AL136638 Homo sapi
17	351.8	13.9	1388	6	AX440923	AX440923 Sequence
18	333.6	13.1	1110	6	BD135990	BD135990 Secretory
c 19	305.8	12.0	132832	9	AC091934	AC091934 Homo sapi
20	300.2	11.8	2502	9	BC002920	BC002920 Homo sapi
21	295.8	11.7	864	6	AX343015	AX343015 Sequence
22	294.2	11.6	368	9	AF174603	AF174603 Homo sapi
23	244.6	9.6	394	6	AX072723	AX072723 Sequence
24	234	9.2	2073	9	BC044244	BC044244 Homo sapi
25	233.2	9.2	852	6	AX247836	AX247836 Sequence
26	233	9.2	389	6	AX072711	AX072711 Sequence
27	223.6	8.8	1995	10	BC024685	BC024685 Mus muscu
28	175	6.9	1998	3	AK116798	AK116798 Ciona int
29	165	6.5	1216	3	AF274013	AF274013 Drosophil
30	164.2	6.5	168318	2	AC024042	AC024042 Homo sapi
c 31	160	6.3	187478	2	AC123758	AC123758 Mus muscu
c 32	160	6.3	215066	2	AC136719	AC136719 Mus muscu
c 33	152.8	6.0	102298	10	AL671671	AL671671 Mouse DNA
34	152.8	6.0	171419	2	AC135142	AC135142 Rattus no
35	152.8	6.0	234117	2	AC130985	AC130985 Rattus no
c 36	152.4	6.0	168318	2	AC024042	AC024042 Homo sapi
37	151.4	6.0	267	6	BD072489	BD072489 Secreted
38	149.2	5.9	123377	9	AC108210	AC108210 Homo sapi
39	149.2	5.9	158785	2	AC027699	AC027699 Homo sapi
40	147.6	5.8	1816	3	BT004914	BT004914 Drosophil
c 41	135.8	5.4	171419	2	AC135142	AC135142 Rattus no
42	131.8	5.2	1661	10	BC003872	BC003872 Mus muscu
43	127.8	5.0	1694	6	BD132536	BD132536 Secreted
44	127.8	5.0	1703	9	AF120266	AF120266 Homo sapi
45	127.8	5.0	1726	9	BC003157	BC003157 Homo sapi

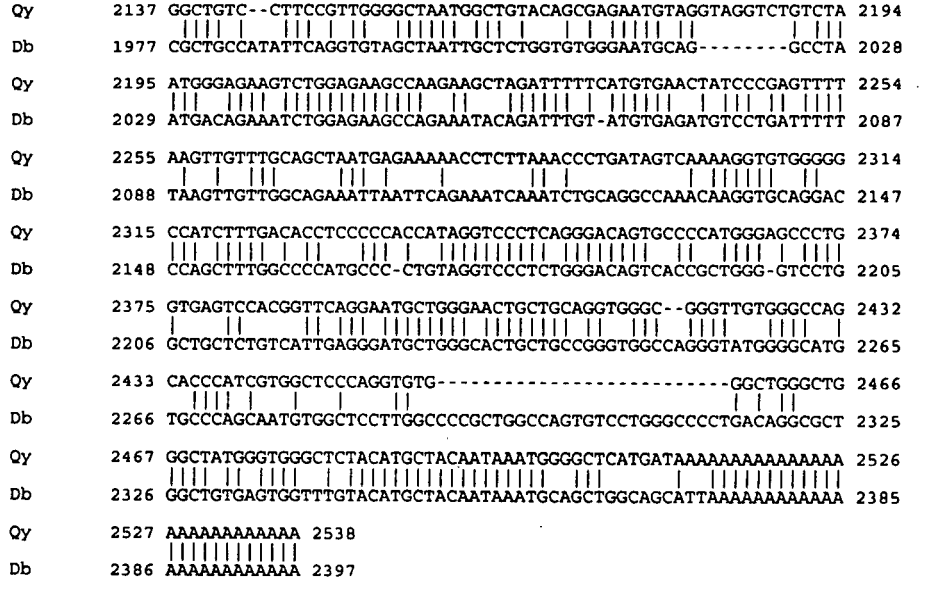
ALIGNMENTS

RESULT 1
BC010405
LOCUS BC010405 2426 bp mRNA linear PRI 10-JUN-2003
DEFINITION Homo sapiens hypothetical protein MGC14859, mRNA (cDNA clone MGC:14859 IMAGE:3621871), complete cds.
ACCESSION BC010405
VERSION BC010405.1 GI:14714540
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2426)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

```
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2426)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 24 Row: k Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction, Similarity but not identity
            to protein.
FEATURES             Location/Qualifiers
     source          1. .2426
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:14859 IMAGE:3621871"
                     /tissue_type="Pancreas, adenocarcinoma"
                     /clone_lib="NIH_MGC_39"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .2426
                     /gene="MGC14859"
                     /note="synonym: FLJ00016"
                     /db_xref="LocusID:89852"
     CDS             61. .852
                     /codon_start=1
                     /product="MGC14859 protein"
                     /protein_id="AAH10405.1"
                     /db_xref="GI:14714541"
                     /db_xref="LocusID:89852"
                     /translation="MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG
                     EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL
                     IFFLELATGILAFVFKDWIRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARG
                     PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDVRLKLELEQQGFI
                     HTKGCVGQFEKWLQDNLIVVAGVFMGIALLQIFGICLAQNLEQME"
BASE COUNT      507 a    662 c    716 g    541 t
```

```
ORIGIN

  Query Match             37.9%;  Score 960.8;  DB 9;  Length 2426;
  Best Local Similarity   69.8%;  Pred. No. 1.5e-205;
  Matches 1768;  Conservative    0;  Mismatches  567;  Indels  197;  Gaps   27;

Qy         64 GAGGGCCGATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCC 123
              |||| | |  |  | |||||||||||||| |||||||||||||| ||| ||||||||||||
Db          6 GAGGCCTGGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGCC 65

Qy        124 GGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTT 183
              |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||
Db         66 CGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTT 125

Qy        184 TGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGC 243
              ||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||||||||
Db        126 TGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGGC 185

Qy        244 CTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGA 303
              |||||||||||||| |||||||| |||||||| |||||||| |||||||||||||| ||
Db        186 CTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTGA 245

Qy        304 CCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTG 363
              ||| |||||||||||||| || |||||||||||||||| |||||||||||||| |||||
Db        246 CCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCTG 305

Qy        364 CATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCT 423
              ||| ||||| |||||||||||||| |||||||||||||||||||||||| |||||||| || ||
Db        306 CATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTCT 365

Qy        424 CATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCG 483
              |||||||||||||||||||||||||||||||| ||||||| || ||||||||||||||||||
Db        366 CATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTCG 425

Qy        484 AGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCT 543
              |||||||||||  ||||||| |||||||||||||||||||| ||||| ||||||||||
Db        426 AGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACCT 485

Qy        544 CCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAA 603
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||| || ||
Db        486 CCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGCGGAGCCCGAGGCCCCAA 545

Qy        604 TGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTG 663
              |||||||||||||||||||| |||||||||||||||||||||| ||||| ||||||||
Db        546 TGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCTG 605

Qy        664 CGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTG 719
              |||||||||||||||||||| |||||||||||||    ||||||||||||||||||||
Db        606 CGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGTG 665

Qy        720 TGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGG 779
              |||||| || |||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db        666 TGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAGG 725

Qy        780 CTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTT 839
              ||| ||||||||||||||||||||||||||||||||||||||||| |||||||| || |||||
Db        726 CTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCTT 785

Qy        840 TGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGA 899
              |||||||||| |||||||||||||||| |||||||||||||||||||| | |
Db        786 CATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGAG----- 840

Qy        900 CATCAAGGCAGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTACTCAAATA 959
                                                                | | |
Db        841 ----------------------------------------------------CAAATGGA 848

Qy        960 AACAAAACCTTGAAAACCACTGGCTTACGCCCACCATCTCAGAGGTTCCATGGGCCGCAG 1019
              |  |  || |||||||||||||||||||||||||||| || ||||| |  |  | || |
Db        849 ATGATGACTTTGAAAACCACTGGCTTACGCCCACCATTTCCGAGGTCCTGTCCACGGCGG 908

Qy       1020 GGCCTCAGCCGTGCCGTCTGCCTGGGGCCCC-AGCCCAGACCCACCCTGCCAACATGTTT 1078
              ||||||||| |  |  |||| |||||||||  |||| | ||||||| ||| |||||||||
Db        909 GGCCTCAGCAGAACTCTCTGACTGGGGCCCCTGGCCCGGCCCCACCCAGCCGACATGTTT 968
```



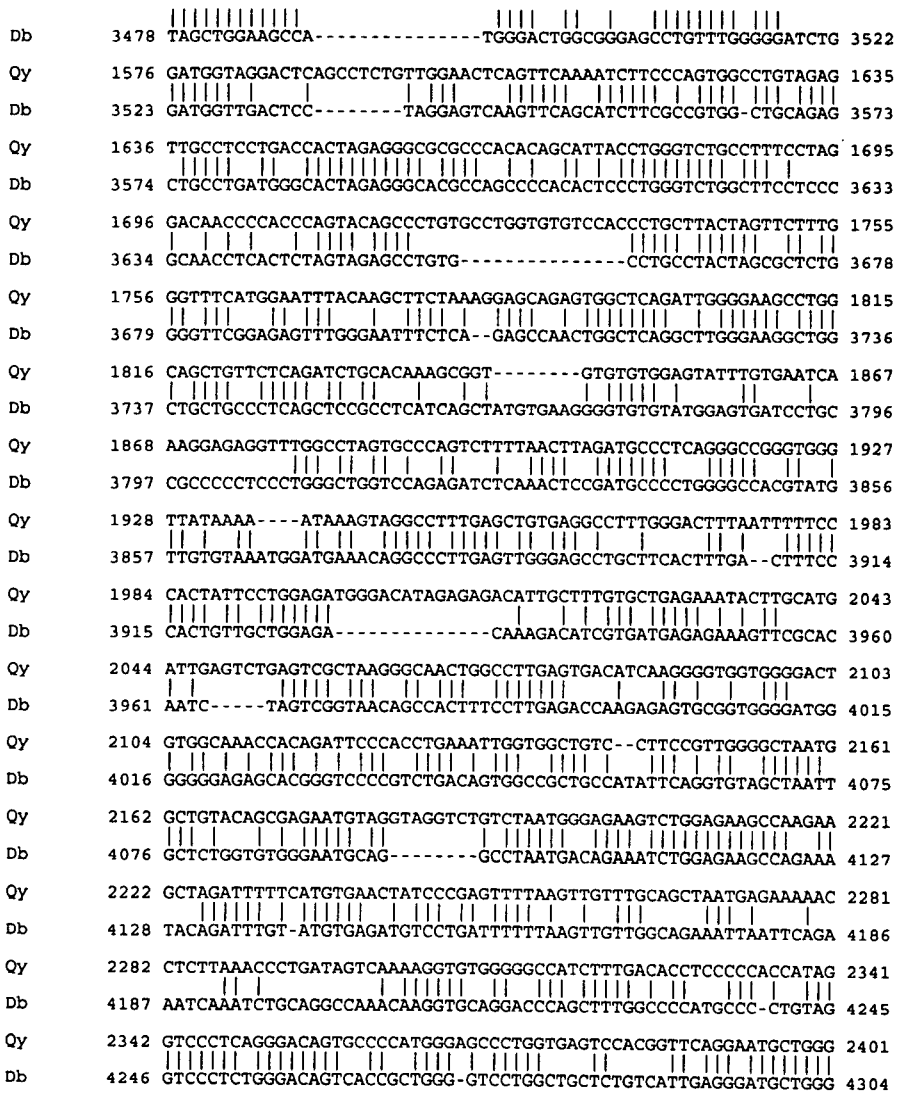


```
RESULT 2
AX420466
LOCUS       AX420466               3184 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 1 from Patent WO0216603.
ACCESSION   AX420466
VERSION     AX420466.1  GI:21524614
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Leiby,K.R.
  TITLE     23228, a human tetraspanin family member and uses thereof
  JOURNAL   Patent: WO 0216603-A 1 28-FEB-2002;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3184
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             168. .980
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD35489.1"
                     /db_xref="GI:21524615"
                     /translation="MPGKHQHFQEPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG
                     EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL
                     IFFLELATGILAFVFKDWIRDQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARG
                     PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDVRLKLELEQQGFI
                     HTKGCVGQFEKWLQDNLIVVAGVFMGIALLQIFGICLAQNLVSDIKAVKANW"
BASE COUNT      675 a    883 c    945 g    681 t
ORIGIN

  Query Match           31.1%;  Score 790.4;  DB 6;  Length 3184;
  Best Local Similarity  92.3%;  Pred. No. 3.7e-167;
  Matches  855;  Conservative    0;  Mismatches   66;  Indels    5;  Gaps    2;

Qy          4 CGCGTCCGGCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTG 63
```

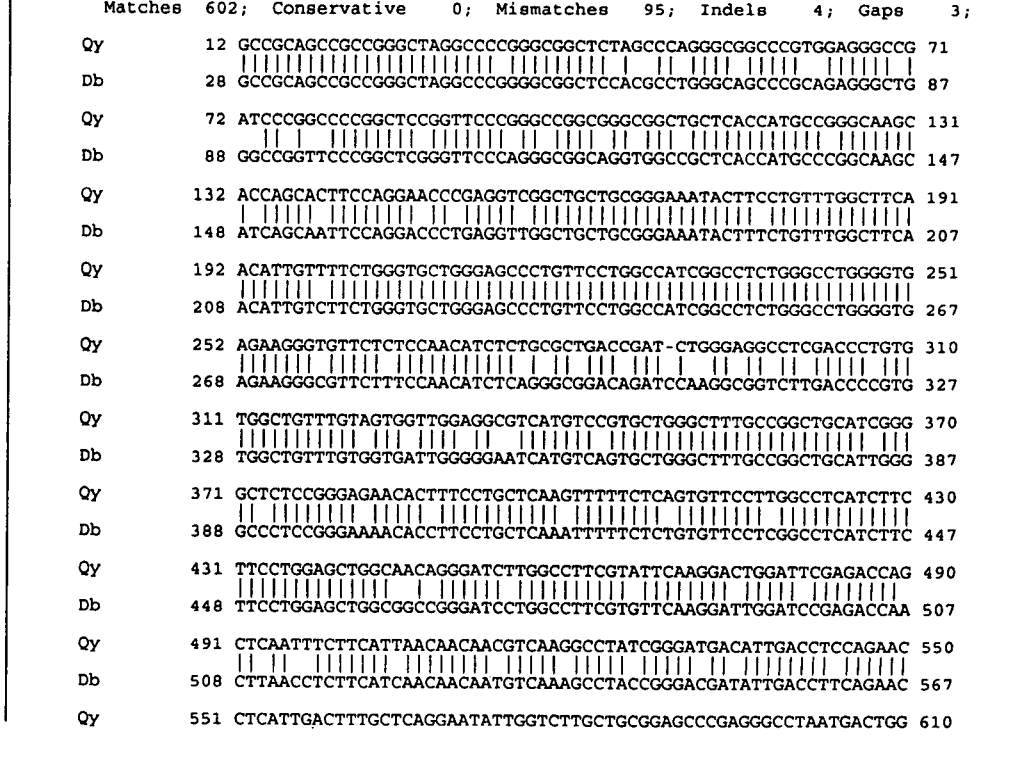
```
Db         52 CGCATGAAGCCGCAGCCGCCCGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGCG 111

Qy         64 GAGGGCCGATCCCGGC-CCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGC 122
              | | || |  || ||| ||||||||||||| |||||||||||| ||| ||||||||||
Db        112 GGGCGCTGGGCCTGGCTCCCGGCTCCGGTTTCCGGGCCGGCGGGTGGCCGCTCACCATGC 171

Qy        123 CGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 182
              | |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||
Db        172 CCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGT 231

Qy        183 TTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGG 242
              |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db        232 TTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTCTGGG 291

Qy        243 CCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCG 302
              |||||||||||||||| |||||||| |||||||| |||||||| |||||||||||||| |
Db        292 CCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCCTTG 351

Qy        303 ACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCT 362
              |||| |||||||||||||| || |||||||||||||||| |||||||||||||| ||||
Db        352 ACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCTGGCT 411

Qy        363 GCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCC 422
              |||| ||||| |||||||||||||| |||||||||||||||||||| |||||||| || |
Db        412 GCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTCGGTC 471

Qy        423 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTC 482
              |||||||||||||||||||||||||||||||||||||| ||||||| || ||||||||||||
Db        472 TCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTC 531

Qy        483 GAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACC 542
              |||||||||||||  ||||||| |||||||||||||||||||| ||||| ||||||||||
Db        532 GAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATTGACC 591

Qy        543 TCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTA 602
              |||||||||||||||||||||||||||||| |||||||||| |||||||||| || |
Db        592 TCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGCCCCA 651

Qy        603 ATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCT 662
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||
Db        652 ATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAGCGCT 711

Qy        663 GCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGT 718
              |||||||||||||||||||| ||||||||||||||    ||||||||||||||||||||
Db        712 GCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACCCAGT 771

Qy        719 GTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778
              ||||||| || |||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db        772 GTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACCAAAG 831

Qy        779 GCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCT 838
              |||| |||||||||||||||||||||||||||||||||||||||| ||||||||| |||| 
Db        832 GCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGAGTCT 891

Qy        839 TTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTG 898
              |  |||||||||| |||||||||||||||| |||||||||||||||||||| |||||||
Db        892 TCATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTGAGTG 951

Qy        899 ACATCAAGGCAGTGAAGGCCAACTGG 924
              |||||||||||||||| |||||||||
Db        952 ACATCAAGGCAGTGAAAGCCAACTGG 977


RESULT 3
BC010346
LOCUS       BC010346               1516 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus RIKEN cDNA 2210021G21 gene, mRNA (cDNA clone MGC:6941
            IMAGE:2811935), complete cds.
ACCESSION   BC010346
VERSION     BC010346.1  GI:16307592
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1516)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 5 Row: f Column: 18.
FEATURES                Location/Qualifiers
     source             1. .1516
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="MGC:6941 IMAGE:2811935"
                        /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
                        old, gross tissue."
                        /clone_lib="NCI_CGAP_Mam5"
                        /lab_host="DH10B"
                        /note="Vector: pCMV-SPORT6"
     gene               1. .1516
                        /gene="2210021G21Rik"
                        /db_xref="LocusID:74257"
                        /db_xref="MGI:1921507"
     CDS                136. .948
                        /codon_start=1
                        /product="RIKEN cDNA 2210021G21"
                        /protein_id="AAH10346.1"
                        /db_xref="GI:16307593"
                        /db_xref="LocusID:74257"
```

```
                  /translation="MPGKHQQFQDPEVGCCGKYFLFGFNIVFWVLGALFLAIGLWAWG
                  EKGVLSNISALTDLGGLDPVWLFVVVGGVMSVLGFAGCIGALRENTFLLKFFSVFLGL
                  IFFLELAAGILAFVFKDWIRDQLNLFINNNVKAYRDDLDLQNLIDFAQEYWSCCGARG
                  PNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVLNTQCGYDIRLKLELEQQGSI
                  YTKGCVGQFEKWLQDNLIVVAGVLVGIALLQIFGLCLAQNLVSDIKAVKANW"
BASE COUNT      277 a    417 c    461 g    361 t
ORIGIN

  Query Match             29.4%;  Score 745.4;  DB 10;  Length 1516;
  Best Local Similarity   89.1%;  Pred. No. 5.2e-157;
  Matches  817;  Conservative    0;  Mismatches   96;  Indels    4;  Gaps    1;

Qy         12 GCCGCAGCCGCCGGGCTAGGCCCCGGGCGGCTCTAGCCCAGGGCGGCCCGTGGAGGGCCG 71
Db         29 GCCGCAGCCACCGGGCTAGGCCCGGGGCTGCTCCAGGCCTGGGCGGCCCGCAGAGGGCTG 88

Qy         72 ATCCCGGCCCCGGCTCCGGTTCCCGGGCCGGCGGGCGGCTGCTCACCATGCCGGGCAAGC 131
Db         89 GGCCGGTCACCGGCTCGGGTTCCCGGGACGGCAGGTGGCCGCTCACCATGCCCGGCAAGC 148

Qy        132 ACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 191
Db        149 ACCAGCAATTCCAGGACCCTGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCA 208

Qy        192 ACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTG 251
Db        209 ACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATTGGCCTCTGGGCCTGGGGTG 268

Qy        252 AGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGT 311
Db        269 AGAAGGGCGTTCTCTCCAACATCTCGGCGCTGACAGATCTGGGCGGTCTTGACCCCGTGT 328

Qy        312 GGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGG 371
Db        329 GGCTGTTTGTGGTGGTGGGGGGAGTCATGTCAGTGTTGGGCTTCGCCGGCTGCATTGGGG 388

Qy        372 CTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCT 431
Db        389 CCCTCCGGGAAAACACCTTCCTGCTCAAGTTCTTCTCTGTGTTCCTCGGCCTCATCTTCT 448

Qy        432 TCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGC 491
Db        449 TCCTGGAGCTGGCCGCGGGGATCCTGGCCTTCGTCTTCAAGGATTGGATCCGAGACCAAC 508

Qy        492 TCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACC 551
Db        509 TTAATCTCTTCATCAACAACAATGTCAAAGCCTACCGGGATGATCTCGACCTTCAGAACC 568

Qy        552 TCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGA 611
Db        569 TTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGTGCCCGAGGGCCCAATGACTGGA 628

Qy        612 ACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGC 671
Db        629 ACCTCAATATCTACTTCAACTGCACTGACTTGAACCCAAGCCGTGAGCGCTGTGGGGTGC 688

Qy        672 CCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATG 727
Db        689 CCTTTTCCTGCTGTGTTAGGGACCCAGCGGAAGATGTCCTCAATACCCAGTGTGGCTATG 748

Qy        728 ATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGG 787
Db        749 ACATCCGACTCAAACTGGAGCTCGAGCAGCAGGGCTCCATCTACACCAAAGGCTGTGTGG 808

Qy        788 GCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCA 847
Db        809 GCCAGTTTGAGAAGTGGCTACAAGACAACCTGATCGTGGTGGCCGGGGTCTTGGTGGGGA 868

Qy        848 TCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGG 907
Db        869 TCGCCCTCCTCCAGATATTTGGCCTCTGCTTGGCCCAGAACCTCGTGAGTGACATCAAGG 928

Qy        908 CAGTGAAGGCCAACTGG 924
Db        929 CGGTGAAGGCCAACTGG 945


RESULT 4
AX420468
LOCUS       AX420468                 813 bp    DNA     linear   PAT 18-JUN-2002
DEFINITION  Sequence 3 from Patent WO0216603.
ACCESSION   AX420468
VERSION     AX420468.1  GI:21524616
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Leiby,K.R.
  TITLE     23228, a human tetraspanin family member and uses thereof
  JOURNAL   Patent: WO 0216603-A 3 28-FEB-2002;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..813
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      151 a    233 c    235 g    194 t
ORIGIN

  Query Match             28.0%;  Score 710.4;  DB 6;  Length 813;
  Best Local Similarity   93.2%;  Pred. No. 4.1e-149;
  Matches  755;  Conservative    0;  Mismatches   51;  Indels    4;  Gaps    1;

Qy        119 ATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTC 178
Db          1 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGGTCGGCTGCTGCGGGAAATACTTC 60

Qy        179 CTGTTTGGCTTCAACATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTC 238
Db         61 CTGTTTGGCTTCAACATTGTCTTCTGGGTGCTGGGAGCCCTGTTCCTGGCTATCGGCCTC 120

Qy        239 TGGGCCTGGGGTGAGAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGC 298
Db        121 TGGGCCTGGGGTGAGAAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 180

Qy        299 CTCGACCCTGTGTGGCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCC 358
Db        181 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCGTCATGTCGGTGCTGGGCTTTGCT 240

Qy        359 GGCTGCATCGGGGCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTT 418
Db        241 GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCCTGCTCAAGTTTTTCTCCGTGTTCCTC 300

Qy        419 GGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGG 478
Db        301 GGTCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGG 360

Qy        479 ATTCGAGACCAGCTCAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATT 538
Db        361 ATTCGAGACCAGCTCAACCTCTTCATCAACAACAACGTCAAGGCCTACCGGGACGACATT 420

Qy        539 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGG 598
Db        421 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGC 480

Qy        599 CCTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAG 658
Db        481 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCAGCCGGGAG 540

Qy        659 CGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACC 714
Db        541 CGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAGGGACCCTGCGGAGGATGTCCTCAACACC 600

Qy        715 CAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACC 774
```

```
                ||||||||||| || |||||||||||||||||||||||||||||||||| ||| ||||||
Db        601 CAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACC 660

Qy        775 AAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGG 834
                |||||||| |||||||||||||||||||||||||||||||||||| |||||||| ||
Db        661 AAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCTGCAGGACAACCTGATTGTGGTGGCGGGA 720

Qy        835 GTCTTTGTGGGCATCGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTG 894
                |||||  |||||||||| |||||||||||||||| |||||||||||||||||||| |||
Db        721 GTCTTCATGGGCATCGCCCTCCTCCAGATCTTTGGCATCTGCCTGGCCCAGAACCTCGTG 780

Qy        895 AGTGACATCAAGGCAGTGAAGGCCAACTGG 924
                |||||||||||||||||||| |||||||||
Db        781 AGTGACATCAAGGCAGTGAAAGCCAACTGG 810


RESULT 5
AK024427
LOCUS       AK024427               4445 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens mRNA for FLJ00016 protein, partial cds.
ACCESSION   AK024427
VERSION     AK024427.1  GI:10440362
KEYWORDS    fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4445)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
  TITLE     The nucleotide sequence of a long cDNA clone isolated from human
            spleen
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 4445)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/NEDO, Tel:81-438-52-3913,
            Fax:81-438-52-3914)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Kazusa DNA Research Institute.
FEATURES             Location/Qualifiers
     source          1. .4445
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="as00016"
                     /tissue_type="spleen"
                     /dev_stage="adult"
                     /note="vector:pBluescriptII SK plus"
     gene            1. .4445
                     /gene="FLJ00016"
     CDS             <2041. .3090
                     /gene="FLJ00016"
                     /note="For this clone, GeneMark analysis triggered an
                     alert for spurious CDS split. The result of GeneMark
                     analysis of the nucleotide sequence of this clone is
                     accessible through http://www.kazusa.or.jp/NEDO.
                     Start codon is not identified."
                     /codon_start=1
                     /product="FLJ00016 protein"
                     /protein_id="BAB15717.1"
                     /db_xref="GI:10440363"
                     /translation="GVLGGGDGVWDDASRHGGNGVARLWGWQDSPPVPALKRGGLCRT
                     SGASVGGLLQTWGPAVSAATGLGPAPASQVVWRKPQSLNQYPCPTPSLFLASLGSGKG
                     ELSVPSLAPLCLHQPGGWAQACGGGEVATHLAVPALSQTALYSCSLGLFPNRVRLLGW
                     AWLTWGSPLPPQWSCCGARGPNDWNLNIYFNCTDLNPSRERCGVPFSCCVRDPAEDVL
                     NTQCGYDVRLKLELEQQGFIHTKGCVGQFEKWLQDNLIVVAGVFMGIALLQIPGICLA
                     QNLVSDIKAVKANWSKWNDDFENHWLTPTISEVLSTAGPQQNSLTGAPGPAPPSRHVF
                                FGLGGLYPEPTFKNW"
BASE COUNT      890 a   1278 c   1321 g    956 t
ORIGIN

  Query Match             22.9%;  Score 580.2;  DB 9;  Length 4445;
  Best Local Similarity   67.0%;  Pred. No. 8.5e-120;
  Matches 1257;  Conservative    0;  Mismatches  483;  Indels  137;  Gaps   25;

Qy        571 ATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCAA 630
                | | |||||||||||||||||||||||||| || |||||||||||||||||||||||| |||||
Db       2553 ACAGTGGTCTTGCTGCGGAGCCCGAGGCCCCAATGACTGGAACCTCAATATCTACTTCAA 2612

Qy        631 CTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCAG 690
                |||||||||||||||||| ||||| |||||||||||||||||||||||||||||| |||||
Db       2613 CTGCACTGACTTGAACCCCAGCCGGGAGCGCTGCGGGGTGCCCTTCTCCTGCTGCGTCAG 2672

Qy        691 GGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGGA 746
                |||||||||    |||||||||||||||||||||||||| || ||||||||||||||||
Db       2673 GGACCCTGCGGAGGATGTCCTCAACACCCAGTGTGGCTACGACGTCCGGCTCAAACTGGA 2732

Qy        747 GCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGCT 806
                |||||||||||||||||| ||| |||||||||||||| |||||||||||||||||||||||
Db       2733 GCTGGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTGGGCCAGTTTGAGAAGTGGCT 2792

Qy        807 GCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCTT 866
                |||||||||||||| |||||||| || |||||  |||||||||| ||||||||||||||
Db       2793 GCAGGACAACCTGATTGTGGTGGCGGGAGTCTTCATGGGCATCGCCCTCCTCCAGATCTT 2852

Qy        867 TGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCCAACTGGAT 926
                ||| |||||||||||||||||| |||||||||||||||||||||| |||||||||
Db       2853 TGGCATCTGCCTGGCCCAGAACCTCGTGAGTGACATCAAGGCAGTGAAAGCCAACTGGAG 2912

Qy        927 CAAACATGATGATGGCTACAAACTACTCAAATAAACAAAACCTTGAAAACCACTGGCTTA 986
                ||||    ||||||                          || ||||||||||||||||||
Db       2913 CAAATGGAATGATG------------------------ACTTTGAAAACCACTGGCTTA 2947

Qy        987 CGCCCACCATCTCAGAGGTTCCATGGGCCGCAGGGCCTCAGCCGTGCCGTCTGCCTGGGG 1046
                |||||||||| || ||||| |  |  | || |||||||||| |  |  |||| ||||||
Db       2948 CGCCCACCATTTCCGAGGTCCTGTCCACGGCGGGGCCTCAGCAGAACTCTCTGACTGGGG 3007

Qy       1047 CCCC-AGCCCAGACCCACCCTGCCAACATGTTTTC-TTGGCCTGGGTAGTACATACGATG 1104
                ||||  |||| | ||||||| ||| |||||||||| |||||||||| ||  ||||  ||
Db       3008 CCCCTGGCCCGGCCCCACCCAGCCGACATGTTTTCTTTGGCCTGGGTGGTTTATACCCTG 3067

Qy       1105 AGCCAACCTTTAAAACTTGGCATATTTCATGTAAAAGTCCAGATCCCCAGCATCTTGTGA 1164
                |||||||||||||||| |||| | |||||| ||||||||||||||| |||| |||  |||
Db       3068 AGCCAACCTTTAAAAATTGGTAGATTTCACATAAAAGTCCAGATCCACAGCTTCTCTTGA 3127

Qy       1165 AGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGCTTCGTCTCCTGGGATGTGCGCTTC 1224
                |||||| ||| | ||| ||  |||||||||  |||| |    |||||||       || |
Db       3128 AGAATGACCACCTGGCTACGCCGGCTCTTCGGTGGCAACACTACCTGGGACACTGCCTCC 3187

Qy       1225 CTGTTCTCTGAGGGACCCACCCTCACCCGTGTCC---------TGCCTGCCTGACCCTGG 1275
                |  || |  |||| |||| |    ||  | | |          ||||  |    || ||
Db       3188 CCAGTCACCAAGGGCCCCAGCTGGCCCGTTCTACTCACCTAAGTGCCGCCTGACCCTTGT 3247

Qy       1276 AGGCTGGGAGCTGGCCTCCTCCACCTCTGCAAGTTTTTCCCCTGCAAATGCTGCAAGGCT 1335
                |  || ||||||||||||  ||||||||||| | || |  ||||||    |  | ||||
Db       3248 ACACTAGGAGCTGGCCTC--CCACCTCTGCAGGGTTATTTCCTGCA----CCTCGAGGCC 3301

Qy       1336 GCTGTGGGCCAAGCCCGGATCGAAGCCTGGAGCGTGAAGAATTGGGGAGGCTGGAGCCTG 1395
                |||| |||||||  |  |  |  ||  | ||  | ||||| | ||| |||||||||| || |
Db       3302 GCTGCGGGCCAATCTGGAGTGAAACACGGGGACCTGAAGGA-TGGAGAGGCTGGACCCCG 3360

Qy       1396 CCCCAAAGAGGCCACAGCCTGGGAAGGGTCTGGCCCTCCTGGGGGCCAAGATGGGTGCCA 1455
                |    ||||||   ||||||||||||||    |||| | ||||||  |    | |||| |  |
Db       3361 CTTTGAAGAGGGTGCAGCCTGGGAAGGG--CGGCCTTGCTGGGGACTGCGGTGGGAG-TA 3417

Qy       1456 CCGTGCCCAGGAGAGTGGCCGGAGGGTGGGATGGAGATCAGGAAGGTTTTGGGCAGGACG 1515
                |||||||||||||| | | | |||||||||| | ||||||   |||||   | || |
Db       3418 GAGTGCCCAGGAGAGGGTCTGAGGGGTGGGATGGGGGTCAGGACAATTTTGCAAAAGAAG 3477

Qy       1516 TAGCTGGAAGCCTGAGCTTGTCACCCATGGGGATGGGGAGAGCCCTGTTTGAGGGCGGCT 1575
```



```
RESULT 6
AX061778
LOCUS       AX061778               1567 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 2 from Patent WO0078948.
ACCESSION   AX061778
VERSION     AX061778.1  GI:12539860
KEYWORDS    .
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Andriamanpandry,C. and Maitre,M.
  TITLE     Cloning, expression and characterisation of a cdna coding for a rat
            brain gamma-hydroxybutyrate (ghb) receptor
  JOURNAL   Patent: WO 0078948-A 2 28-DEC-2000;
            Universite Louis Pasteur de Strasbourg (FR)
FEATURES             Location/Qualifiers
     source          1. .1567
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
     CDS             22. .1560
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC25029.1"
                     /db_xref="GI:12539861"
                     /translation="MRAAAAGLGPGRLHAWAARRGLGRFPARVPRAAGGRSPCPASIS
                     NSRTLRLAAAGNTFCLASTLSSGCWEPCSWPSASGPGVRRAFFPTSQGGQIQGGLDPV
                     WLFVVIGGIMSVLGFAGCIGALRENTFLLKFFSVFLGLIFFLELAAGILAFVFKDWIR
                     DQLNLFINNNVKAYRDDIDLQNLIDFAQEYWSCCGARGPNDWNLNIRTSTALTSNPSR
                     ERCGVPFFCWVRTLRKTSQYPCGYTSAQTELEHKIHLHQSWWPFEKWLKKPDGWPGLG
                     AIASFKMGIAGPNPRSTSRQERPTGKLPGVMATCLAHRAVVGASKDALPHSQWQGLWD
                     VCYDLSPRLCEHGTQEATEAGLGLNWGCTGAGLGLFTTELCVWGVHMCVCVCVCVCVR
                     VCLCLCVRVRGMHVCALVSTPGVSTPLLVRWWPFQSFKGDGARRVGHVPASPSLLWDV
                     SLCGLGACCLRPLHIHHDLEPAWSSPWPQCHSLEMGPRILSVSLSRLPLRTHTWQDGY
                     ERLN"
BASE COUNT      281 a    436 c    480 g    370 t
ORIGIN
```

```
                 || || |||||||||||||||||| |||||||||| |||||||||||||| |||||||||
Db        568 CTTATCGACTTTGCTCAGGAATACTGGTCTTGCTGTGGAGCCCGAGGGCCCAATGACTGG 627

Qy        611 AACCTCAATATC--TATTTCAACTGCACTGACTT-GAACCCGAGCCGAGAGCGCTGCGGG 667
              |||||||| |||   | ||||||||||||||||   ||||| ||||| |||||||| |||
Db        628 AACCTCAACATCCGGACTTCAACTGCACTGACTTCAAACCCAAGCCGCGAGCGCTGTGGG 687

Qy        668 GTGCCCTTCTCCTGCTGTGTCAGGGACCCTGCGATGTCCTC 708
              |||||||||| |||||| || ||| ||    || | | ||
Db        688 GTGCCCTTCTTCTGCTGGGTAAGGACCCTGCGGAAGACGTC 728


RESULT 7
BC041304
LOCUS       BC041304                1685 bp    mRNA    linear   VRT 14-JAN-2003
DEFINITION  Xenopus laevis, Similar to transmembrane 4 superfamily member 9,
            clone IMAGE:4683897, mRNA, partial cds.
ACCESSION   BC041304
VERSION     BC041304.1  GI:27735442
KEYWORDS    .
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1685)
  AUTHORS   Klein,S. and Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-2002) National Institutes of Health, Xenopus Gene
            Collection (XGC), National Institute of Child Health and Human
            Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
            20892-7510, USA
  REMARK    NIH-MGC Project
COMMENT     Contact: XGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Igor Dawid
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 94 Row: g Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .1685
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4683897"
                     /tissue_type="Embryo, stage 31/32, Xenopus"
                     /clone_lib="NICHD_XGC_Emb4"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             <1. .926
                     /codon_start=3
                     /product="Similar to transmembrane 4 superfamily member 9"
                     /protein_id="AAH41304.1"
                     /db_xref="GI:27735443"
                     /translation="FGSTRSVGLIYCFHLAPSRAVSPDPPRCQPGNLPSCSGSMSGKH
                     YKGPEVSCCIKYFIFGFNVIFWLLGLTFLGVGLWAWSEKGVLSNISSITDLGGFDPVW
                     LFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIFFLELTAGVLAFVFKDWIKD
                     QLQFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNLNIYFNCTDSNASRER
                     CGVPFSCCTKDPAEDVINTQCGYDVRQKPELDQQETIHTKGCVPQFEKWLQDNLTIVA
                     GVFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      456 a    332 c    383 g    514 t
ORIGIN
```

```
  Query Match           18.3%;  Score 465.2;  DB 5;  Length 1685;
  Best Local Similarity  73.7%;  Pred. No. 7.1e-94;
  Matches  607;  Conservative    0;  Mismatches  213;  Indels    4;  Gaps    1;

Qy        133 CCAGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAA 192
              | |||||| | ||| |||||||||| ||||||||   || |||||| | || ||||||||
Db        128 CAAGCACTACAAGGGACCCGAGGTCAGCTGCTGCATCAAGTACTTCATATTCGGCTTCAA 187

Qy        193 CATTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGA 252
              |  | ||||| |  |||||     ||| || |   | || || ||||| ||| ||||
Db        188 TGTCATCTTCTGGCTCTTGGGACTTACGTTTCTTGGAGTTGGTCTGTGGGCATGGAGTGA 247

Qy        253 GAAGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTG 312
              |||||||| || |||||||| ||   | | || ||||| || ||  | || || |||||
Db        248 AAAGGGTGTGCTTTCCAACATTTCCTCTATCACAGATCTTGGGGGTTTTGATCCAGTGTG 307

Qy        313 GCTGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGC 372
              ||||||| | || || ||||| |||||||   | |||||||||||| ||||| || |||||
Db        308 GCTGTTTCTGGTAGTCGGAGGGGTCATGTTTATCCTGGGCTTTGCAGGCTGTATTGGGGC 367

Qy        373 TCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTT 432
              ||||| |||||||| |||||| | || ||||| || ||||| |||||  | || |||||
Db        368 ACTCCGTGAGAACACCTTCCTGTTGAAAATTTTTTTCTGTGTTTCTTGGAATTATTTTCTT 427

Qy        433 CCTGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCT 492
              ||||||||  | | ||  |  | |||||||| |||| |||||||||  |||||| ||
Db        428 TCTGGAGCTCACGGCGGGAGTTCTTGCCTTCGTGTTCAAAGACTGGATTAAAGACCAACT 487

Qy        493 CAATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCT 552
              | ||||| |||||||||||| |||  ||||  | |||||||| ||| | ||||||||
Db        488 GCAGTTCTTTATTAACAACAACATCAGAGCCTACAGAGATGACATCGACTTACAGAACCT 547

Qy        553 CATTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAA 612
              ||||||||  | ||||||||||||   |||||||| ||     ||  || |||| ||||
Db        548 TATTGACTTCACACAGGAATATTGGCAGTGCTGCGGGGCGTTCGGAGCTGATGATTGGAA 607

Qy        613 CCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCC 672
              | || || |||||||||||||| |||| ||| |||||||||||| | || |||||||||
Db        608 TTTAAACATTTATTTCAACTGCACCGACTCTAACGCGAGCCGAGAGAGATGTGGGGTGCC 667

Qy        673 CTTCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGA 728
              || |||||||||     | ||||||||     |||||  | |||||  ||||||||||||||
Db        668 ATTTTCCTGCTGCACAAAAGACCCTGCTGAGGATGTAATAAACACACAGTGTGGCTATGA 727

Qy        729 TGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGG 788
              |||| | |  || |  || ||||| || ||||    | || ||||| ||||||||||
Db        728 TGTCAGACAAAAGCCTGAACTGGATCAACAGGAGACAATTCACACAAAAGGCTGTGTCCC 787

Qy        789 CCAGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCAT 848
              |||||||||| |||||||||| ||||| | | | ||||| ||||| ||  | || ||
Db        788 ACAGTTTGAGAAATGGCTGCAGGATAACCTTACCATTGTGGCAGGGGTGTTCATTGGTAT 847

Qy        849 CGCTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGC 908
              || ||  | ||||| ||||| |||||  | || |||||  | ||||| |||||  | ||
Db        848 TGCGCTTTTACAGATATTTGGAATCTGTTTAGCGCAGAATTTGGTGAGCGACATTGAAGC 907

Qy        909 AGTGAAGGCCAACTGGATCAAACATGATGATGGCTACAAACTAC 952
              ||    |||| ||||   ||  || ||  ||| |||||| | |
Db        908 CGTCCGAGCCAGCTGGTAGAATTATTATCCTGGATACAAAGTGC 951


RESULT 8
AF121344
LOCUS       AF121344                3175 bp    mRNA    linear   ROD 01-FEB-2000
DEFINITION  Mus musculus tetraspanin Tspan-5 (Tspan5) mRNA, complete cds.
ACCESSION   AF121344
VERSION     AF121344.1  GI:6841032
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 3175)
  AUTHORS   Garcia-Frigola,C., de Lecea,L. and Soriano,E.
  TITLE     Mouse Tspan-5 cDNA cloning
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3175)
  AUTHORS   Garcia-Frigola,C., de Lecea,L. and Soriano,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1999) Dept. of Animal and Plant Cell Biology,
            University of Barcelona, Av. Diagonal 645, Barcelona 08028, Spain
FEATURES             Location/Qualifiers
     source          1. .3175
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Swiss Webster/NIH"
                     /db_xref="taxon:10090"
     gene            1. .3175
                     /gene="Tspan5"
     CDS             388. .1194
                     /gene="Tspan5"
                     /note="member of transmembrane 4 superfamily"
                     /codon_start=1
                     /product="tetraspanin Tspan-5"
                     /protein_id="AAF28869.1"
                     /db_xref="GI:6841033"
                     /translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
                     GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIF
                     FLELTAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGAD
                     DWNLNIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYT
                     KGCVPQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      667 a    811 c    840 g    857 t
ORIGIN

  Query Match             18.0%;  Score 456.4;  DB 10;  Length 3175;
  Best Local Similarity   73.9%;  Pred. No. 6.7e-92;
  Matches  594;  Conservative    0;  Mismatches  206;  Indels    4;  Gaps    1;

Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        398 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATTTTTGGCTTCAATG 457

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||     ||| || |  ||||| || ||||| |||  ||| |
Db        458 TCATATTTTGGTTTTTGGGAATAACGTTTCTTGGAATCGGACTGTGGGCGTGGAATGAAA 517

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | |||||| ||||||||||||||   |  | ||||| || || ||| | ||||| |||||||
Db        518 AAGGTGTCCTCTCCAACATCTCGTCCATCACCGACCTCGGTGGCTTTGACCCAGTGTGGC 577

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | | ||||| ||||| || |||||||| || |
Db        578 TTTTCCTTGTGGTGGGAGGAGTGATGTTCATTCTGGGGTTTGCAGGGTGCATCGGAGCAC 637

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCC 434
              | ||||| ||||| || || |||||||||||| || |||||||| ||   | || |||||||
Db        638 TTCGGGAAAACACCTTTCTTCTCAAGTTTTTTTCTGTGTTCCTGGGGATTATTTTCTTCC 697

Qy        435 TGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCA 494
              |||| ||  |  | ||| | ||||| || || ||||| |||||||   |||||||||
Db        698 TGGAACTCACTGCTGGGGTGTTGGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTGT 757

Qy        495 ATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCA 554
              ||||||| |||||||||||| |||  |||||  | |||||||||| || ||||||||||
Db        758 ATTTCTTTATTAACAACAACATCAGAGCCTACAGAGATGACATTGATCTACAGAACCTCA 817

Qy        555 TTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACC 614
              | |||||  | ||||||||||||   ||||| || ||    ||  || |||| |||||||
Db        818 TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAACC 877

Qy        615 TCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCT 674
              | ||||| || ||||| ||||| || |  ||  | |||||||||| ||||| |||| |
Db        878 TAAATATTTACTTCAATTGCACAGATTCCAATGCAAGCCGAGAGCGATGCGGTGTGCCAT 937

Qy        675 TCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATG 730
              | |||||||    |  ||||| ||    |||||| ||||||| ||||||||||||||||
Db        938 TTTCCTGCTGCACTAAAGACCCCGCGGAAGATGTCATCAACACTCAGTGTGGCTATGATG 997

Qy        731 TCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCC 790
              | |||  ||||  ||  | || || |||       ||  |||| ||||||||||||  ||
Db        998 CCAGGCAGAAACCAGAAGTTGACCAACAGATTGTAATCTACACAAAAGGCTGTGTGCCCC 1057

Qy        791 AGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCG 850
              |||||||||||||||| ||||||||  | | | | ||||||||  | ||  | |||| |
Db        1058 AGTTTGAGAAGTGGCTACAGGACAATTTAACCATCGTGGCTGGTATTTTCATAGGCATTG 1117

Qy        851 CTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAG 910
              | | || |||| |||| || |||||||| |||||  | ||||||||||  | || |
Db        1118 CATTGCTACAGATTTTTGGGATATGCCTGGCACAGAATTTGGTGAGTGACATTGAAGCTG 1177

Qy        911 TGAAGGCCAACTGGATCAAACATG 934
              | | ||| | ||||   |  | ||
Db        1178 TCAGGGCTAGCTGGTAGACCCCTG 1201


RESULT 9
AF065389
LOCUS       AF065389               1405 bp    mRNA    linear   PRI 28-APR-2000
DEFINITION  Homo sapiens tetraspan NET-4 mRNA, complete cds.
ACCESSION   AF065389
VERSION     AF065389.1  GI:3152702
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1405)
  AUTHORS   Serru,V., Dessen,P., Boucheix,C. and Rubinstein,E.
  TITLE     Sequence and expression of seven new tetraspans
  JOURNAL   Biochim. Biophys. Acta 1478 (1), 159-163 (2000)
  MEDLINE   20185353
   PUBMED   10719184
REFERENCE   2  (bases 1 to 1405)
  AUTHORS   Rubinstein,E., Serru,V. and Boucheix,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1998) INSERM U268, 14 av Paul Vaillant Couturier,
            Villejuif 94807, France
FEATURES             Location/Qualifiers
     source          1. .1405
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE ID 219547"
     CDS             348. .1154
                     /note="similar to Tspan-5; TM4SF"
                     /codon_start=1
                     /product="tetraspan NET-4"
                     /protein_id="AAC17120.1"
                     /db_xref="GI:3152703"
                     /translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
                     GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIF
                     FLELTAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGAD
                     DWNLNIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYT
                     KGCVPQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      309 a    357 c    387 g    352 t
ORIGIN

  Query Match             17.4%;  Score 441.6;  DB 9;  Length 1405;
  Best Local Similarity   73.2%;  Pred. No. 1.5e-88;
  Matches  581;  Conservative    0;  Mismatches  209;  Indels    4;  Gaps    1;

Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        358 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 417

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||     || || |  || || || ||||| |||  ||| |
Db        418 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 477
```

```
Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | || ||||| |||||||||||| |  | |||||||| || ||| | ||||| || ||||
Db        478 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 537

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | |  |||| ||||| || ||||| || || |
Db        538 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGCGC 597

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCC 434
              | ||||| |||||||||| |||||||||| || |||||||| ||  | || |||||||
Db        598 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATTATTTTCTTCC 657

Qy        435 TGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCA 494
              |||||||  |  | ||  |  | || || || ||||| ||||||||   |||||||||
Db        658 TGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTGT 717

Qy        495 ATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCA 554
              ||||||| || |||||||||| |||  || ||||||||||||||||  | || |||||||
Db        718 ATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTCA 777

Qy        555 TTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACC 614
              | |||||  | ||||||||||||   ||||| || ||    ||  || |||| |||||||
Db        778 TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAACC 837

Qy        615 TCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCT 674
              | ||||| || ||||| ||||| || |  ||  | || |||||||| || || || || |
Db        838 TAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCAT 897

Qy        675 TCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATG 730
              |||||||||    | || || ||    |||||| |||||| ||||||||||||||||
Db        898 TCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGATG 957

Qy        731 TCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCC 790
               | |||  ||||  ||  | || ||||||       ||  |||| ||||||||||||  ||
Db        958 CCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCCC 1017

Qy        791 AGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCG 850
              |||||||||||||| |||||||||  | | | | || |||||  | ||  | |||||  |
Db       1018 AGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATTG 1077

Qy        851 CTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAG 910
              |  | || ||||| ||||| || ||||||||||||||  | || || || ||| | || |
Db       1078 CATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCTG 1137

Qy        911 TGAAGGCCAACTGG 924
              | | ||| | ||||
Db       1138 TCAGGGCGAGCTGG 1151
```

```
RESULT 10
BC009704
LOCUS       BC009704                1416 bp    mRNA    linear   PRI 13-JAN-2003
DEFINITION  Homo sapiens, tetraspan 5, clone MGC:9300 IMAGE:3895933, mRNA,
            complete cds.
ACCESSION   BC009704
VERSION     BC009704.1  GI:16307230
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1416)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 14 Row: h Column: 4
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21264582.
FEATURES             Location/Qualifiers
     source          1. .1416
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:9300 IMAGE:3895933"
                     /tissue_type="Pancreas, epithelioid carcinoma"
                     /clone_lib="NIH_MGC_70"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             334. .1140
                     /codon_start=1
                     /product="tetraspan 5"
                     /protein_id="AAH09704.1"
                     /db_xref="GI:16307231"
                     /db_xref="LocusID:10098"
                     /translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
                     GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLLKFFSVFLGIIF
                     FLELTAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGAD
                     DWNLNIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYT
                     KGCVPQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT      325 a    354 c    385 g    352 t
ORIGIN
```

```
  Query Match             17.4%;  Score 441.6;  DB 9;  Length 1416;
  Best Local Similarity   73.2%;  Pred. No. 1.5e-88;
  Matches  581;  Conservative    0;  Mismatches  209;  Indels    4;  Gaps    1;

Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        344 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 403

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              | | || ||| |  |||||      || || |  || || || ||||| |||  ||| |
Db        404 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 463

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | || ||||| |||||||||||| |  | |||||||| || ||| | ||||| || ||||
Db        464 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 523

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | |  |||| ||||| || ||||| || || |
Db        524 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGCGC 583

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCC 434
              | ||||| |||||||||| |||||||||| || |||||||| ||  | || |||||||
Db        584 TACGGGAAAACACTTTCCTTCTCAAGTTTTTTTCTGTGTTCCTGGGAATTATTTTCTTCC 643

Qy        435 TGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCA 494
              |||||||  |  | ||  |  | || || || ||||| ||||||||   |||||||||
Db        644 TGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTGT 703

Qy        495 ATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCA 554
              ||||||| || |||||||||| |||  || ||||||||||||||||  | || |||||||
Db        704 ATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTCA 763

Qy        555 TTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACC 614
              | |||||  | ||||||||||||   ||||| || ||    ||  || |||| |||||||
Db        764 TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAACC 823
```

```
Qy        615 TCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCT 674
              | ||||| || ||||| ||||| || |   ||   | || |||||||| || || || || |
Db        824 TAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCAT 883

Qy        675 TCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATG 730
              ||||||||||     |  || || ||    |||||| ||||||| ||||||||||||||||
Db        884 TCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGATG 943

Qy        731 TCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCC 790
              | |||  ||||  ||  | || ||||||      ||  |||| ||||||||||||  ||
Db        944 CCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCCC 1003

Qy        791 AGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCG 850
              |||||||||||||| ||||||||||  | | | | || |||||  | ||  | |||||
Db       1004 AGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATTG 1063

Qy        851 CTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAG 910
              |  | || ||||| ||||| || ||||||||||||||  | || || || ||| | || |
Db       1064 CATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCTG 1123

Qy        911 TGAAGGCCAACTGG 924
              | | ||| | ||||
Db       1124 TCAGGGCGAGCTGG 1137
```

RESULT 11
AF053455
LOCUS AF053455 1408 bp DNA linear PRI 03-NOV-1998
DEFINITION Homo sapiens tetraspan TM4SF (TSPAN-5) gene, complete cds.
ACCESSION AF053455
VERSION AF053455.1 GI:2995864
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1408)
AUTHORS Todd,S.C., Doctor,V.S. and Levy,S.
TITLE Sequences and expression of six new members of the tetraspanin/TM4SF family
JOURNAL Biochim. Biophys. Acta 1399 (1), 101-104 (1998)
MEDLINE 98390278
PUBMED 9714763
REFERENCE 2 (bases 1 to 1408)
AUTHORS Todd,S.C., Doctor,V.S. and Levy,S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1998) Medicine, Stanford, 300 Pasteur Dr, Stanford, CA 94305, USA
FEATURES Location/Qualifiers
source 1. .1408
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
gene 352. .1146
/gene="TSPAN-5"
CDS 352. .1146
/gene="TSPAN-5"
/note="transmembrane 4 superfamily"
/codon_start=1
/product="tetraspan TM4SF"
/protein_id="AAC69712.1"
/db_xref="GI:2995865"
/translation="MSGKHYKGPEVSCCIKYFIFGFNVIFWFLGITFLGIGLWAWNEK
GVLSNISSITDLGGFDPVWLFLVVGGVMFILGFAGCIGALRENTFLSVFLGIIFFLEL
TAGVLAFVFKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFTQEYWQCCGAFGADDWNL
NIYFNCTDSNASRERCGVPFSCCTKDPAEDVINTQCGYDARQKPEVDQQIVIYTKGCV
PQFEKWLQDNLTIVAGIFIGIALLQIFGICLAQNLVSDIEAVRASW"
BASE COUNT 309 a 354 c 384 g 345 t 16 others
ORIGIN

Query Match 16.1%; Score 409.2; DB 9; Length 1408;
Best Local Similarity 71.8%; Pred. No. 3e-81;

Matches 570; Conservative 0; Mismatches 208; Indels 16; Gaps 2;

```
Qy        135 AGCACTTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACA 194
              |||||| | |||  || || ||| | || |||   ||||||||| | ||||||||||
Db        362 AGCACTACAAGGGTCCTGAAGTCAGTTGTTGCATCAAATACTTCATATTTGGCTTCAATG 421

Qy        195 TTGTTTTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGA 254
              |  | || ||| |  |||||       || || |  || || || ||||| |||  ||| |
Db        422 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAATTGGACTGTGGGCATGGAATGAAA 481

Qy        255 AGGGTGTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGC 314
              | || ||||| |||||||||||| |  | |||||||| || ||| | ||||| || ||||
Db        482 AAGGAGTTCTGTCCAACATCTCTTCCATCACCGATCTCGGCGGCTTTGACCCAGTTTGGC 541

Qy        315 TGTTTGTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTC 374
              | ||  | ||||| ||||| || |||| | |  |||| ||||| || ||||| || || |
Db        542 TCTTCCTTGTGGTGGGAGGAGTGATGTTCATTTTGGGATTTGCAGGGTGCATTGGAGCGC 601

Qy        375 TCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCC 434
              | ||||| ||||||||||             | || |||||||| ||  | || |||||||
Db        602 TACGGGAAAACACTTTCC-----------TTTCTGTGTTCCTGGGAATTATTTTCTTCC 649

Qy        435 TGGAGCTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCA 494
              |||||||  |  | ||  |  | || || || ||||| ||||||||   |||||||||
Db        650 TGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTTCAAAGACTGGATCAAAGACCAGCTGT 709

Qy        495 ATTTCTTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCA 554
              ||||||| || |||||||||| |||  || ||||||||||||||||||  | || |||||||
Db        710 ATTTCTTTATAAACAACAACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTCA 769

Qy        555 TTGACTTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACC 614
              | |||||  | ||||||||||||    ||||| || ||      ||  || |||| |||||||
Db        770 TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTGGAACC 829

Qy        615 TCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCT 674
              | ||||| || ||||| ||||| || |   ||   | || |||||||| || || || || |
Db        830 TAAATATTTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGTGGCGTTCCAT 889

Qy        675 TCTCCTGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATG 730
              ||||||||||     |  || || ||    |||||| ||||||| ||||||||||||||||
Db        890 TCTCCTGCTGCACTAAAGATCCCGCAGAAGATGTCATCAACACTCAGTGTGGCTATGATG 949

Qy        731 TCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCC 790
              | |||  ||||  ||  | || ||||||      ||  |||| ||||||||||||  ||
Db        950 CCAGGCAAAAACCAGAAGTTGACCAGCAGATTGTAATCTACACGAAAGGCTGTGTGCCCC 1009

Qy        791 AGTTTGAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCG 850
              |||||||||||||| ||||||||||  | | | | || |||||  | ||  | |||||
Db       1010 AGTTTGAGAAGTGGTTGCAGGACAATTTAACCATCGTTGCTGGTATTTTCATAGGCATTG 1069

Qy        851 CTCTCCTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAG 910
              |  | || ||||| ||||| || ||||||||||||||  | || || || ||| | || |
Db       1070 CATTGCTGCAGATATTTGGGATATGCCTGGCCCAGAATTTGGTTAGCGATATCGAAGCTG 1129

Qy        911 TGAAGGCCAACTGG 924
              | | ||| | ||||
Db       1130 TCAGGGCGAGCTGG 1143
```

RESULT 12
BC025568
LOCUS BC025568 2428 bp mRNA linear ROD 16-APR-2003
DEFINITION Mus musculus DNA segment, Chr 14, ERATO Doi 226, expressed, mRNA (cDNA clone MGC:36595 IMAGE:5322531), complete cds.
ACCESSION BC025568
VERSION BC025568.1 GI:19343883
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2428)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 2428)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 56 Row: j Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.

FEATURES Location/Qualifiers

source 1. .2428
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:36595 IMAGE:5322531"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

gene 1. .2428
/gene="D14Ertd226e"
/note="synonym: MGC36554"
/db_xref="LocusID:52588"
/db_xref="MGI:1196325"

CDS 64. .876
/codon_start=1
/product="tetraspanin similar to TM4SF9"
/protein_id="AAH25568.1"
/db_xref="GI:19343884"
/db_xref="LocusID:52588"
/translation="MHYYRYSNAEVSCWYKYLLFSYNIVFWLAGVVFLGVGLWAWSEK
GVLSDLTKVTRLHGIDPVVLVLMVGVVMFTLGFAGCVGALRENICLLKFFCGAIVLIF
FLELAVAVLAFLFQDWVRDRFREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPE
DWDLNVYFNCSGASYSREKCGVPFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDEFIFT
KGCIQALEGWLPRNIYIVAGVFIAISLLQIFGIFLARTLISDIEAVKAGHHF"

BASE COUNT 517 a 640 c 624 g 647 t
ORIGIN

Query Match 14.5%; Score 368.8; DB 10; Length 2428;
Best Local Similarity 67.9%; Pred. No. 3.7e-72;
Matches 531; Conservative 0; Mismatches 247; Indels 4; Gaps 1;

```
Qy        140 TTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTT 199
              | |  | |  |||||||| |||||||     || ||| | || ||  ||| ||| || ||
Db         79 TACTCGAACGCCGAGGTCAGCTGCTGGTACAAGTACCTGCTCTTCAGCTACAATATCGTC 138

Qy        200 TTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGT 259
              || ||| ||   |||| | ||||| |   |||| || ||||| ||| | || ||||||
Db        139 TTTTGGCTGGCTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGT 198

Qy        260 GTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTT 319
              || || ||| || || |  | |||||   ||   ||  ||||||| ||   |||| |
Db        199 GTGCTGTCCGACCTCACCAAGGTGACCCGGTTGCATGGAATCGACCCCGTCGTGCTGGTC 258

Qy        320 GTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGG 379
              |  |||| || |  || |||| |  ||||| || || |||||  ||||||| |||||
Db        259 TTGATGGTGGGCGTGGTGATGTTCACACTGGGATTCGCAGGCTGTGTCGGGGCCCTCCGA 318

Qy        380 GAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAG 439
              |||||||  | ||||||||||||||||  | |  | |||  ||||||||||||||||||
Db        319 GAGAACATCTGCCTGCTCAAGTTTTTCTGTGGGGCCATTGTGCTCATCTTCTTCCTGGAA 378

Qy        440 CTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTC 499
              |||||   |  | ||||||||  ||||| | ||||||  |  ||||||  | ||
Db        379 CTGGCCGTGGCCGTGTTGGCCTTTTTATTCCAAGACTGGGTGAGAGACCGGTTCCGGGAA 438

Qy        500 TTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGAC 559
              ||| |  | | |||| |||| |||||||||||||||| ||||| ||||||||||||||||
Db        439 TTCTTCGAGAGCAACATCAAGTCCTATCGGGATGACATCGACCTGCAGAACCTCATTGAC 498

Qy        560 TTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAAT 619
              |    |||| ||  |       |||||||| ||    || ||| | |||||| ||||||||
Db        499 TCCCTTCAGAAAGCTAATCAGTGCTGCGGGGCTTACGGCCCTGAAGACTGGGACCTCAAT 558

Qy        620 ATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCC 679
              |||| |||||||||| ||     | |  |||||||||   || ||||| |||||||||
Db        559 GTCTACTTCAACTGCAGTGGTGCCAGCTACAGCCGAGAGAAATGTGGGGTACCCTTCTCC 618

Qy        680 TGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATGTCCGG 735
              |||||||||    || |||||      | ||| | ||||| ||||||||||||||||||||
Db        619 TGCTGTGTGCCAGATCCTGCACAAAAAGTCGTGAACACACAGTGTGGCTATGATGTCCGG 678

Qy        736 CTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTT 795
              |  | ||| ||    ||  | | |  | |||    || ||||| ||  |         |
Db        679 ATTCAGCTGAAGAGCAAGTGGGATGAGTTCATCTTTACAAAAGGATGCATCCAGGCTCTG 738

Qy        796 GAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTC 855
              ||   |||||||      ||| |   | | |||||||| |||||  | | |||| | ||
Db        739 GAAGGCTGGCTGCCCAGGAACATCTACATTGTGGCTGGTGTCTTCATTGCCATCTCACTG 798

Qy        856 CTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAG 915
              || ||||| |||| |||| ||||||    || |||  |     |||||| ||||||||||
Db        799 CTGCAGATTTTTGGCATCTTCCTGGCGAGGACCCTGATCTCAGACATCGAGGCAGTGAAG 858

Qy        916 GC 917
              ||
Db        859 GC 860
```

```
RESULT 13
BC024611
LOCUS       BC024611               2498 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus DNA segment, Chr 14, ERATO Doi 226, expressed, mRNA
            (cDNA clone MGC:28503 IMAGE:4188261), complete cds.
ACCESSION   BC024611
VERSION     BC024611.1  GI:19354053
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2498)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2498)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 37 Row: e Column: 12
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2498
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:28503 IMAGE:4188261"
                     /tissue_type="Salivary gland, 10 week old female mouse"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .2498
                     /gene="D14Ertd226e"
                     /note="synonym: MGC36554"
                     /db_xref="LocusID:52588"
                     /db_xref="MGI:1196325"
     CDS             55. .867
                     /codon_start=1
                     /product="tetraspanin similar to TM4SF9"
                     /protein_id="AAH24611.1"
                     /db_xref="GI:19354054"
                     /db_xref="LocusID:52588"
                     /translation="MHYYRYSNAEVSCWYKYLLFSYNIVFWLAGVVFLGVGLWAWSEK
                     GVLSDLTKVTRLHGIDPVVLVLMVGVVMFTLGFAGCVGALRENICLLKFFCGAIVLIF
                     FLELAVAVLAFLFQDWVRDRFREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPE
                     DWDLNVYFNCSGASYSREKCGVPFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDEFIFT
                     KGCIQALEGWLPRNIYIVAGVFIAISLLQIFGIFLARTLISDIEAVKAGHHF"
BASE COUNT      542 a    644 c    631 g    681 t
ORIGIN

  Query Match           14.5%;  Score 368.8;  DB 10;  Length 2498;
  Best Local Similarity 67.9%;  Pred. No. 3.7e-72;
  Matches  531;  Conservative    0;  Mismatches  247;  Indels    4;  Gaps    1;

Qy        140 TTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTT 199
              | |  | |  |||||||| |||||||    || ||| | || ||  ||| ||| || ||
Db         70 TACTCGAACGCCGAGGTCAGCTGCTGGTACAAGTACCTGCTCTTCAGCTACAATATCGTC 129

Qy        200 TTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGT 259
              || ||| ||   ||||   | ||||| |   |||| || ||||| ||| | || ||||||
Db        130 TTTTGGCTGGCTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGT 189

Qy        260 GTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTT 319
              || || ||| || || |   | |||||    ||   ||  ||||||| ||   |||| |
Db        190 GTGCTGTCCGACCTCACCAAGGTGACCCGGTTGCATGGAATCGACCCCGTCGTGCTGGTC 249

Qy        320 GTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGG 379
              |  |||| || |  || |||| |   ||||| || || |||||  ||||||| |||||
Db        250 TTGATGGTGGGCGTGGTGATGTTCACACTGGGATTCGCAGGCTGTGTCGGGGCCCTCCGA 309

Qy        380 GAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAG 439
              |||||||  | ||||||||||||||||  | |  | |||  ||||||||||||||||||
Db        310 GAGAACATCTGCCTGCTCAAGTTTTTCTGTGGGGCCATTGTGCTCATCTTCTTCCTGGAA 369

Qy        440 CTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTC 499
              |||||    |   | ||||||| ||||| | |||||| |  |||||| | ||
Db        370 CTGGCCGTGGCCGTGTTGGCCTTTTTATTCCAAGACTGGGTGAGAGACCGGTTCCGGGAA 429

Qy        500 TTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGAC 559
              ||| |  | | |||| ||||| ||||||||||||||| ||||| ||||||||||||||
Db        430 TTCTTCGAGAGCAACATCAAGTCCTATCGGGATGACATCGACCTGCAGAACCTCATTGAC 489

Qy        560 TTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAAT 619
              |    |||| ||  |      |||||||| ||    || ||| | |||||| ||||||||
Db        490 TCCCTTCAGAAAGCTAATCAGTGCTGCGGGGCTTACGGCCCTGAAGACTGGGACCTCAAT 549

Qy        620 ATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCC 679
              |||| |||||||||| ||    | |  |||||||||  || ||||| |||||||||
Db        550 GTCTACTTCAACTGCAGTGGTGCCAGCTACAGCCGAGAGAAATGTGGGGTACCCTTCTCC 609

Qy        680 TGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATGTCCGG 735
              ||||||||    || |||||     | ||| | ||||| ||||||||||||||||||||
Db        610 TGCTGTGTGCCAGATCCTGCACAAAAAGTCGTGAACACACAGTGTGGCTATGATGTCCGG 669

Qy        736 CTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTT 795
              |  | ||| ||    ||  | | |  | |||    || ||||| ||  |       |
Db        670 ATTCAGCTGAAGAGCAAGTGGGATGAGTTCATCTTTACAAAAGGATGCATCCAGGCTCTG 729

Qy        796 GAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTC 855
              ||    |||||||     ||| |   | | |||||||| |||||  | | |||| | ||
Db        730 GAAGGCTGGCTGCCCAGGAACATCTACATTGTGGCTGGTGTCTTCATTGCCATCTCACTG 789
```

```
Qy        856 CTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAG 915
              || ||||| ||||| |||| ||||||   || |||  |    |||||| ||||||||||
Db        790 CTGCAGATTTTTGGCATCTTCCTGGCGAGGACCCTGATCTCAGACATCGAGGCAGTGAAG 849

Qy        916 GC 917
              ||
Db        850 GC 851
```

RESULT 14
BC026574
LOCUS BC026574 2500 bp mRNA linear ROD 16-APR-2003
DEFINITION Mus musculus DNA segment, Chr 14, ERATO Doi 226, expressed, mRNA (cDNA clone MGC:36554 IMAGE:4954874), complete cds.
ACCESSION BC026574
VERSION BC026574.1 GI:20072352
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2500)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2500)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 59 Row: a Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.
FEATURES Location/Qualifiers
source 1. .2500
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:36554 IMAGE:4954874"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
gene 1. .2500
/gene="D14Ertd226e"
/note="synonym: MGC36554"
/db_xref="LocusID:52588"
/db_xref="MGI:1196325"
CDS 50. .862
/codon_start=1
/product="tetraspanin similar to TM4SF9"
/protein_id="AAH26574.1"
/db_xref="GI:20072353"
/db_xref="LocusID:52588"
/translation="MHYYRYSNAEVSCWYKYLLFSYNIVFWLAGVVFLGVGLWAWSEK GVLSDLTKVTRLHGIDPVVLVLMVGVVMFTLGFAGCVGALRENICLLKFFCGAIVLIF FLELAVAVLAFLFQDWVRDRFREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPE DWDLNVYFNCSGASYSREKCGVPFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDEFIFT KGCIQALEGWLPRNIYIVAGVFIAISLLQIFGIFLARTLISDIEAVKAGHHF"
BASE COUNT 545 a 643 c 631 g 681 t
ORIGIN

Query Match 14.5%; Score 368.8; DB 10; Length 2500;
Best Local Similarity 67.9%; Pred. No. 3.7e-72;
Matches 531; Conservative 0; Mismatches 247; Indels 4; Gaps 1;

```
Qy        140 TTCCAGGAACCCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTT 199
              | |  | |  |||||||| |||||||     || ||| | || ||  ||| ||| || ||
Db         65 TACTCGAACGCCGAGGTCAGCTGCTGGTACAAGTACCTGCTCTTCAGCTACAATATCGTC 124

Qy        200 TTCTGGGTGCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGT 259
              || ||| ||   ||||   | ||||| |   |||| || ||||| ||| | || ||||||
Db        125 TTTTGGCTGGCTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGT 184

Qy        260 GTTCTCTCCAACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTT 319
              || || ||| || || |   | |||||    ||   ||  |||||||  ||   |||| |
Db        185 GTGCTGTCCGACCTCACCAAGGTGACCCGGTTGCATGGAATCGACCCCGTCGTGCTGGTC 244

Qy        320 GTAGTGGTTGGAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGG 379
              | |||| || |  || |||| |   ||||| || || |||||  |||||||  |||||
Db        245 TTGATGGTGGGCGTGGTGATGTTCACACTGGGATTCGCAGGCTGTGTCGGGGCCCTCCGA 304

Qy        380 GAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAG 439
              |||||||  | ||||||||||||||||  | |  | |||  ||||||||||||||||||
Db        305 GAGAACATCTGCCTGCTCAAGTTTTTCTGTGGGGCCATTGTGCTCATCTTCTTCCTGGAA 364

Qy        440 CTGGCAACAGGGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTC 499
              |||||    |   | ||||||||  ||||| | ||||||  |  ||||||  | ||
Db        365 CTGGCCGTGGCCGTGTTGGCCTTTTTATTCCAAGACTGGGTGAGAGACCGGTTCCGGGAA 424

Qy        500 TTCATTAACAACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGAC 559
              ||| |  | | |||| ||||| ||||||||||||||| ||||| ||||||||||||||
Db        425 TTCTTCGAGAGCAACATCAAGTCCTATCGGGATGACATCGACCTGCAGAACCTCATTGAC 484

Qy        560 TTTGCTCAGGAATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAAT 619
              |    |||| ||  |      |||||||  ||    || ||| | |||||| ||||||||
Db        485 TCCCTTCAGAAAGCTAATCAGTGCTGCGGGGCTTACGGCCCTGAAGACTGGGACCTCAAT 544

Qy        620 ATCTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCC 679
              |||| |||||||||| ||       | |  |||||||||    || ||||| |||||||||
Db        545 GTCTACTTCAACTGCAGTGGTGCCAGCTACAGCCGAGAGAAATGTGGGGTACCCTTCTCC 604

Qy        680 TGCTGTGTCAGGGACCCTGC----GATGTCCTCAACACCCAGTGTGGCTATGATGTCCGG 735
              ||||||||    || |||||      | ||| | ||||| ||||||||||||||||||||
Db        605 TGCTGTGTGCCAGATCCTGCACAAAAAGTCGTGAACACACAGTGTGGCTATGATGTCCGG 664
```

```
Qy        736 CTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTT 795
Db        665 ATTCAGCTGAAGAGCAAGTGGGATGAGTTCATCTTTACAAAAGGATGCATCCAGGCTCTG 724

Qy        796 GAGAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTC 855
Db        725 GAAGGCTGGCTGCCCAGGAACATCTACATTGTGGCTGGTGTCTTCATTGCCATCTCACTG 784

Qy        856 CTCCAGATCTTTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAG 915
Db        785 CTGCAGATTTTTGGCATCTTCCTGGCGAGGACCCTGATCTCAGACATCGAGGCAGTGAAG 844

Qy        916 GC 917
Db        845 GC 846


RESULT 15
AF311903
LOCUS       AF311903                2553 bp    mRNA    linear   PRI 16-AUG-2002
DEFINITION  Homo sapiens DC-TM4F2 precursor, mRNA, complete cds.
ACCESSION   AF311903
VERSION     AF311903.1  GI:22266721
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2553)
  AUTHORS   Zhang,W., Li,N., Wan,T. and Cao,X.
  TITLE     Identification of novel membrane proteins
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2553)
  AUTHORS   Zhang,W., Li,N., Wan,T. and Cao,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2000) Department of Immunology, Second Military
            Medical University & Shanghai Brilliance Biotechnology Institute,
            800 Xiangyin Rd., Shanghai 200433, P.R.China
FEATURES             Location/Qualifiers
     source          1. .2553
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             73. .885
                     /codon_start=1
                     /product="DC-TM4F2 precursor"
                     /protein_id="AAM94899.1"
                     /db_xref="GI:22266722"
                     /translation="MHYYRYSNAKVSCWYKYLLFSYNIIFWLAGVVFLGVGLWAWSEK
                     GVLSDLTKVTRMHGIDPVVLVLMVGVVMFTLGFAGCVGALRENICLLNFFCGTIVLIF
                     FLELAVAVLAFLFQDWVRDRFREFFESNIKSYRDDIDLQNLIDSLQKANQCCGAYGPE
                     DWDLNVYFNCSGASYSREKCGVPFSCCVPDPAQKVVNTQCGYDVRIQLKSKWDESIFT
                     KGCIQALESWLPRNIYIVAGVFIAISLLQIFGIFLARTLISDIEAVKAGHHF"
BASE COUNT      505 a    724 c    680 g    644 t
ORIGIN

  Query Match           13.9%;  Score 353.4;  DB 9;  Length 2553;
  Best Local Similarity  67.0%;  Pred. No. 1.1e-68;
  Matches  518;  Conservative    0;  Mismatches  251;  Indels    4;  Gaps    1;

Qy        150 CCGAGGTCGGCTGCTGCGGGAAATACTTCCTGTTTGGCTTCAACATTGTTTTCTGGGTGC 209
Db         98 CCAAGGTCAGCTGCTGGTACAAGTACCTCCTTTTCAGCTACAACATCATCTTCTGGTTGG 157

Qy        210 TGGGAGCCCTGTTCCTGGCCATCGGCCTCTGGGCCTGGGGTGAGAAGGGTGTTCTCTCCA 269
Db        158 CTGGAGTTGTCTTCCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAAGGGTGTGCTGTCCG 217

Qy        270 ACATCTCTGCGCTGACCGATCTGGGAGGCCTCGACCCTGTGTGGCTGTTTGTAGTGGTTG 329
Db        218 ACCTCACCAAAGTGACCCGGATGCATGGAATCGACCCTGTGGTGCTGGTCCTGATGGTGG 277

Qy        330 GAGGCGTCATGTCCGTGCTGGGCTTTGCCGGCTGCATCGGGGCTCTCCGGGAGAACACTT 389
Db        278 GCGTGGTGATGTTCACCCTGGGGTTCGCCGGCTGCGTGGGGGCTCTGCGGGAGAATATCT 337

Qy        390 TCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTCATCTTCTTCCTGGAGCTGGCAACAG 449
Db        338 GCTTGCTCAACTTTTTCTGTGGCACCATCGTGCTCATCTTCTTCCTGGAGCTGGCTGTGG 397

Qy        450 GGATCTTGGCCTTCGTATTCAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACA 509
Db        398 CCGTGCTGGCCTTCCTGTTCCAGGACTGGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGA 457

Qy        510 ACAACGTCAAGGCCTATCGGGATGACATTGACCTCCAGAACCTCATTGACTTTGCTCAGG 569
Db        458 GCAACATCAAGTCCTACCGGGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGA 517

Qy        570 AATATTGGTCTTGCTGCGGAGCCCGAGGGCCTAATGACTGGAACCTCAATATCTATTTCA 629
Db        518 AAGCTAACCAGTGCTGTGGCGCATATGGCCCTGAAGACTGGGACCTCAACGTCTACTTCA 577

Qy        630 ACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGCGGGGTGCCCTTCTCCTGCTGTGTCA 689
Db        578 ATTGCAGCGGTGCCAGCTACAGCCGAGAGAAGTGCGGGGTCCCCTTCTCCTGCTGCGTGC 637

Qy        690 GGGACCCTGCG----ATGTCCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGG 745
Db        638 CAGATCCTGCGCAAAAAGTTGTGAACACACAGTGTGGATATGATGTCAGGATTCAGCTGA 697

Qy        746 AGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGCCAGTTTGAGAAGTGGC 805
Db        698 AGAGCAAGTGGGATGAGTCCATCTTCACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGC 757

Qy        806 TGCAGGACAACCTGATCGTGGTGGCTGGGGTCTTTGTGGGCATCGCTCTCCTCCAGATCT 865
Db        758 TCCCGCGGAACATTTACATTGTGGCTGGCGTCTTCATCGCCATCTCGCTGTTGCAGATAT 817

Qy        866 TTGGTATCTGCCTGGCCCAGAACCTTGTGAGTGACATCAAGGCAGTGAAGGCC 918
Db        818 TTGGCATCTTCCTGGCAAGGACGCTGATCTCAGACATCGAGGCAGTGAAGGCC 870


Search completed: November 21, 2003, 21:30:42
Job time : 9172 secs
```